

# The Comparative Benefits Modeling Project



## A Framework for Cost-Utility Analysis of Government Health Care Programs



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Public Health Service



# **The Comparative Benefits Modeling Project**

## **A Framework for Cost-Utility Analysis of Government Health Care Programs**

**Mark S. Kamlet, Principal Investigator**

A report to the Office of Disease Prevention and Health Promotion, Public Health Service, U.S. Department of Health and Human Services, in cooperation with the Foundation for Health Services Research.



**COMPARATIVE BENEFITS MODELING PROJECT  
EXPERT REVIEW PANEL MEMBERS**

Gordon DeFrieze, Ph.D.  
University of North Carolina at  
Chapel Hill

John Drabek, Ph.D.  
Department of Health and Human Services

Pennifer Erickson  
National Center for Health Statistics

J. Michael Fitzmaurice, Ph.D.  
National Center for Health Services Research

William T. Friedwald, M.D.  
National Institutes of Health

Philip Jacobs, D.Phil.  
University of Alberta

James M. Kaple, Ph.D.  
Alcohol, Drug Abuse and Mental Health  
Administration

Barbara H. Kehrler, Ph.D.  
Henry J. Kaiser Family Foundation

Jeffery P. Koplan, M.D.  
Centers for Disease Control

Judith R. Lave, Ph.D.  
University of Pittsburgh

Joseph Lipscomb, Ph.D.  
Duke University

Bryan Luce, Ph.D.  
Batelle MEDTAP

George Scheiber, Ph.D.  
Health Care Financing Administration

Katharine Bauer Sommers  
Institute of Medicine

Donald M. Steinwachs, Ph.D.  
The Johns Hopkins University

Judith Wagner, Ph.D.  
Office of Technology Assessment

John E. Ware, Jr., Ph.D.  
New England Medical Center

Milton Weinstein, Ph.D.  
Harvard University

Harry R. Woltman, Ph.D. (Retired)  
Food and Drug Administration

## **STAFF**

### **THE OFFICE OF DISEASE PREVENTION AND HEALTH PROMOTION**

**James Harrell**  
Office of Disease Prevention and Health  
Promotion

**Marthe Gold, M.D., M.P.H.**  
Office of Disease Prevention and Health  
Promotion

**David Adock, M.D., M.P.H.**  
Luther L. Terry Preventive Medicine Fellow

### **THE FOUNDATION FOR HEALTH SERVICES RESEARCH**

**Alice S. Hersh, M.M.H.S.**  
The Foundation for Health Services Branch

**Laika K. Tamny, M.P.A.**  
The Foundation for Health Services Research

### **PROJECT STAFF**

**Mark S. Kamlet, Ph.D.**  
Carnegie Mellon University

## ACKNOWLEDGEMENT

The Comparative Benefits Modeling Project has attempted to present the principles and tools of cost-effectiveness analysis in health care in a manner useful for nonspecialists while, at the same time, not shrinking from addressing many of the key controversies in the field. This task is ambitious and challenging. While the field of cost-utility analysis continues to mature, at present it lacks a firm foundation of the sort that characterizes such tools as social cost-benefit analysis or decision analysis. Recognizing this, I hope it is not long before this monograph is seen as being, if not outdated, at least badly in need of augmentation to incorporate the burgeoning empirical and theoretical work in cost-utility analysis that I have no doubt will be produced in the coming years.

To the degree to which this monograph is even partially successful in meeting its charge, thanks are due to the help of the unusually talented and patient set of individuals who have been associated with the project. Dr. Michael McGinnis, Director of the Office of Disease Prevention and Health Promotion, and Alice Hersch and Laika Tamney from the Foundation for Health Services Research provided key guidance and support throughout. In three separate meetings, the Expert Panel provided great help in clarifying useful (and, of equal value, not so useful) directions for this work. Joseph Lipscomb went beyond the call of duty in helping to lead one of these expert panel meetings. Milton Weinstein and Thomas McGuire helped me refine my thinking on several important points. Marcia Wade, Steven Klepper, and Marilyn Quadrel helped greatly early in the project. Very special thanks go to David Adcock and Marthe Gold serving successively in the Office of Disease Prevention and Health Promotion. David Adcock was invaluable during the first half of the project, gently prodding me to put together a text that was approachable by someone who had not spent a graduate career studying the Savage axioms of subjective expected utility theory and patiently hearing out my arguments as to why some discussion of foundational issues could not be avoided. Without David I am sure that the first drafts of this work would never have been possible. It is fully to the credit of Marthe Gold that the project subsequently moved forward during a period when my own schedule made this difficult. Her suggestions regarding the substance and style of the text were innumerable and invaluable. With all this help, I surely must take full credit for any remaining lack of clarity or execution in the text.

Mark S. Kamlet  
1992



## TABLE OF CONTENTS

|                                                                                  |             |
|----------------------------------------------------------------------------------|-------------|
| <b>FOREWORD</b> .....                                                            | <b>xi</b>   |
| <b>PREFACE</b> .....                                                             | <b>xiii</b> |
| <b>INTRODUCTION</b> .....                                                        | <b>xv</b>   |
| <b>PART I: THE MECHANICS OF CUA IN HEALTH CARE</b>                               |             |
| <b>1. An Initial Definition of CUA in Health Care</b>                            |             |
| Introduction: Cost-Utility Analysis .....                                        | 3           |
| The Basic Steps of CUA .....                                                     | 3           |
| Narrowing the Definition: What "Cost-Utility Effectiveness" Is Not .....         | 4           |
| Concluding Comment .....                                                         | 8           |
| <b>2. Applying CUA to Evaluate Mutually Compatible Health Interventions</b>      |             |
| Introduction: Application of CUA .....                                           | 9           |
| Maximizing Health Benefits Subject to a Budget Constraint .....                  | 9           |
| <b>3. Applying CUA to Evaluate Mutually Exclusive Health Interventions</b>       |             |
| Introduction .....                                                               | 15          |
| A Basic Algorithm .....                                                          | 15          |
| A Numerical Example .....                                                        | 17          |
| An Alternative Set of Algorithms .....                                           | 19          |
| <b>4. Applying CUA to Choosing the Magnitude of Health Interventions</b>         |             |
| Introduction .....                                                               | 23          |
| The Condition $MC/MB = R_{\text{Threshold}}$ .....                               | 23          |
| $MC/MB = R_{\text{Threshold}}$ Is a Necessary but Not Sufficient Condition ..... | 25          |
| A General Algorithm .....                                                        | 29          |
| If "Nothing" Is Best, the Next Best May Not Be "Next to Nothing" .....           | 29          |
| Measuring Scale .....                                                            | 30          |
| Mixing the "Which" and the "How Much" Questions .....                            | 30          |
| <b>5. Other Considerations In Applying the Mechanics of CUA</b>                  |             |
| Introduction .....                                                               | 33          |
| What if the Numerator and Denominator in the C-E Ratio Are Not Both              |             |
| Positive? .....                                                                  | 33          |
| There Is Nothing Special About the Status Quo .....                              | 35          |

## **PART II: NORMATIVE UNDERPINNINGS TO CUA**

### **6. CUA-I: The Individual Perspective and the Basics of VNM-Utility Theory**

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| Introduction . . . . .                                                  | 43 |
| Basics of Utility Theory: The Unexpected Utility Theorem . . . . .      | 43 |
| An Illustration . . . . .                                               | 45 |
| Linking VNM-Utility Theory to CUA . . . . .                             | 46 |
| Addressing the Interpretive Questions . . . . .                         | 47 |
| Some Basic Multiattribute Utility Notation . . . . .                    | 48 |
| Can Utility Be Attached to Life and Death? . . . . .                    | 49 |
| Implications of VNM-Utility Theory for Health-Status Measures . . . . . | 51 |
| Relaxing the Budget-Constraint Assumption . . . . .                     | 51 |
| Additional Notes on Relaxing the Budget-Constraint Assumption . . . . . | 53 |

### **7. CUA-II: The Social Perspective**

|                                                               |    |
|---------------------------------------------------------------|----|
| Introduction . . . . .                                        | 55 |
| A Constitutional Convention on Health Care . . . . .          | 55 |
| Linking Constitutional Choice to VNM Utility . . . . .        | 58 |
| Linking Constitutional Choice to CUA . . . . .                | 59 |
| Merging Interpersonal and Intrapersonal Comparisons . . . . . | 61 |
| $R_{\text{Threshold}}$ and CUA-II . . . . .                   | 61 |
| Some Concluding Comments . . . . .                            | 63 |

### **8. The Relationship of CUA to CBA, DA, and COIA**

|                                         |    |
|-----------------------------------------|----|
| Introduction . . . . .                  | 65 |
| Cost-of-Illness Analysis . . . . .      | 65 |
| CUA and Cost-Benefit Analysis . . . . . | 66 |
| CUA and Decision Analysis . . . . .     | 75 |

## **PART III: ISSUES IN MAKING CUA OPERATIONAL**

### **9. Complications When There Is More to Life than Health**

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| Introduction . . . . .                                                     | 83 |
| There Is More to Life than Health . . . . .                                | 83 |
| A Fundamental Equivalence Result for CUA-I . . . . .                       | 86 |
| Some Caveats to the Fundamental Equivalence Result . . . . .               | 88 |
| Complications Moving from CUA-I to CUA-II . . . . .                        | 90 |
| What Should Be in the Denominator of a Cost-Effectiveness Ratio? . . . . . | 90 |

### **10. QALYs as an Outcome Measure for CUA**

|                                                                      |     |
|----------------------------------------------------------------------|-----|
| Introduction . . . . .                                               | 99  |
| Use of QALYs in CUA: The Basic Framework . . . . .                   | 99  |
| Extending the Basic Framework . . . . .                              | 101 |
| QALYs and VNM Utility . . . . .                                      | 104 |
| What if Individuals Are "Risk Averse" over Length of Life? . . . . . | 106 |
| Some Other Caveats with QALYs . . . . .                              | 108 |
| Eliciting QALYs . . . . .                                            | 110 |

|                                                                   |         |
|-------------------------------------------------------------------|---------|
| <b>11. Additional Considerations: Uncertainty and Discounting</b> |         |
| Introduction . . . . .                                            | 117     |
| Dealing with Uncertainty in CUA . . . . .                         | 117     |
| Discounting . . . . .                                             | 122     |
| <br><b>CONCLUSION</b> . . . . .                                   | <br>129 |
| <br><b>REFERENCES</b> . . . . .                                   | <br>131 |



## FOREWORD

A fundamental obligation of the public health community is to establish a rational perspective on health interventions. Whether it is an individual's perspective on how to control daily risks to health or society's perspective on how to spend tax dollars for health, the wisest decisions are likely to emerge from the broadest and most balanced perspectives. Health policymakers, in particular, must be able to view the entire range of interventions, preventive as well as therapeutic, and to assess and compare health returns on the Nation's investment in these interventions.

The problem, of course, is how to compare different kinds of returns on dissimilar interventions. In addressing this central question, the Comparative Benefits Modeling Project proceeded on four general assumptions:

1. Quantitative methods can aid decision making related to allocation of resources.
2. Cost-effectiveness analysis (CEA) and its derivative, cost-utility analysis (CUA), have significant potential for this purpose, if carefully conducted and presented in an appropriate (e.g., comparative) context.
3. Many uncertainties are inherent in the conduct of most analyses of cost-effectiveness; dealing with these uncertainties could be facilitated through development of an appropriate model.
4. Application of any such model on a regular basis and across a range of issues would lead to gradual improvement in methods and an improved perspective on the relative returns to society of various health programs.

Accordingly, the Project set out to review the state of the art of CEA and CUA and to develop an appropriate model. To this end, it brought together an expert panel of economists, physicians, public health specialists, and others concerned with health care resource allocation. Although the panel found a high degree of controversy in this young field, it also found a basis for a unified framework through which CUA can be applied to health interventions.

This monograph provides an overview of CUA procedures that are widely accepted, and, where there is controversy, attempts to clarify the issues of debate. It also examines the ethical norms that underlie CUA. Although the monograph focuses on evaluation in the context of governmental resource allocation, the framework, and variants thereof, may be useful to decision makers in the private and not-for-profit sectors as well.

The monograph was prepared for practitioners who wish to apply CUA in a scientific way across a wide range of studies. Its purpose is to provide a clearer perspective from which to view competing claims for health care resources and, ultimately, to aid in decisions concerning allocation of those resources. Because CUA is a young and evolving field, it is hoped that the framework presented here also will help form the basis for greater cumulativeness in research and greater comparability among studies of CEA and CUA in health care.

I would like to express my sincere appreciation to the author, the members of the expert panel, the project staff, and the Foundation for Health Services Research. An unusual degree of cooperation was obtained from all participants in this project, and that is clearly reflected in the quality of the monograph.

J. Michael McGinnis, M.D.  
Deputy Assistant Secretary for Health  
and  
Director, Office of Disease Prevention  
and Health Promotion

## **PREFACE**

It is increasingly important that the tools we have for aiding policymakers and administrators in making informed choices are actually applied. One of the more frequently discussed methods involves the assessment of cost and effectiveness (utility) across alternative programs. Ideally, public programs should be evaluated periodically in terms of costs and effectiveness, and this information should be used in deciding whether to initiate, continue or terminate programs. This volume provides a state-of-the-art review of these methods to aid sophisticated uses in their application. It is invaluable as a reference source.

The work of Dr. Kamlet represents a significant contribution in many respects. One of the more important is that it provides a framework for comparing health programs that have very different objectives, including health promotion, prevention of disease, acute treatment and long term rehabilitation. Although this is an ambitious objective, it is exceptionally important and has been accomplished by the author.

The purpose of the Foundation for Health Services Research is to promote professional education and to advance the field of health services research. In particular, we seek to support scholarly works, such as this project, that integrate and summarize what has been learned. Priority is given to works that are expected to have direct relevance for health policymakers and administrators.

The Foundation wants to express its appreciation to the author, to the expert review panel, and to the Office of Health Promotion and Disease Prevention whose sponsorship and leadership made this project possible.

Donald Steinwachs, Ph.D.  
Past President  
Foundation for Health Services Research



## INTRODUCTION

Cost-effectiveness analysis (CEA) in health policy, and its close relative, cost-utility analysis (CUA), are methods of summarizing information on the relationship between the resources expended on a health intervention and the health outcomes that result from the intervention. Resource use is measured in monetary terms while outcomes are measured in numerical terms, as a change in health or functional status. By focusing on resource use and outcome measures, CEA/CUA is designed to assist policymakers in resource allocation decisions.

The roots of CEA/CUA in health care extend back some 20 years. During this time a growing group of practitioners and researchers have made substantial progress in defining CEA as a powerful methodology. There has been steady movement toward an accepted set of procedures and interpretations for CEA/CUA. But this movement is far from completed and the field is immature in many respects. Important controversies about how to conduct CEA/CUA and how to interpret the results remain. The goal of the Comparative Benefits Modeling Project has been to try to develop a theoretical framework and model for CUA. The extent to which this succeeds will only be known with the careful application of the tenets developed here to real-world problems in the prioritization of resources.

### **Resource Allocation in Health Care: The Importance of Systematic Analysis**

Concern with the costs and associated health benefits of different health interventions, and with assessing these costs and benefits from a public policy perspective, is especially timely. Health care consumes 12 percent of the nation's gross national product, up from less than 6 percent 20 years ago. This represents a per-capita level of health care expenditures that is some 40 percent greater than that of any other developed country. There are various reasons for this pronounced increase in U.S. health care expenditures, some of which augur further increases. One underlying force has been the changing age profile of the U.S. population, attributable to the maturing of the post-World War II baby boom generation and the increased life span of older Americans. Coupled with the establishment and growth of Government programs such as Medicare and Medicaid, this has led to increased demand for medical care.

At the same time, powerful but expensive new health-care options have emerged on a regular basis, ranging from coronary artery by-pass surgery—rare 20 years ago but decidedly common today—to organ replacement by transplant. The \$50 X-ray now is augmented by a \$400 CAT scan which, in turn, is augmented by a \$1,000 nuclear magnetic resonance image.

Over at least the past decade, as health care demand and health care options have grown, so too has concern for controlling health care costs. Pressures for cost control stem in part from private sector employers and insurance companies, which have seen health-insurance costs and health expenditures rise at rates that far exceed the general inflation rate. The pressure is equally intense in Federal, State, and local governments. At the Federal level, for instance, the last decade has seen prospective payment schemes replace cost reimbursement methods in Medicare payments, a major structural change that was driven in large part by the need to control costs. Such changes notwithstanding, Medicare and Medicaid remain among the fastest growing components of the Federal budget and, increasingly, are central targets in attempts to reduce the Federal deficit. More than ever before, health policy must take into account the economic, as well as the clinical, dimensions of health care. CEA and CUA, along with cost-benefit analysis and decision analysis, are at the heart of this new orientation. The emphasis on economic accountability in health care is likely to grow even

stronger, as many of the underlying forces that cause increases in medical demand and medical costs continue to hold sway.

Not only has interest in CEA/CUA in health care increased in the recent past, but there is growing recognition of the importance of disease prevention and health promotion as critical components of national health policy.

The U.S. Public Health Service (PHS), for example, has evinced a growing interest in this field. In 1979, it published a general blueprint for the field, *Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention*, and in 1980, guided by that blueprint, it published 226 specific disease prevention and health promotion objectives for the Nation. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives*, released in 1990, builds on and advances the agenda for this area of endeavor.

Despite this growing interest in health promotion and disease prevention, there are many important policy issues to be resolved. Scholars such as Norman Daniels write, "Many critics of our health-care system have argued that it is biased in favor of high-technology acute care and that it ignores preventive measures, which are of comparable or greater importance."<sup>1</sup> Bruce MacLaury, President of the Brookings Institution, casts the argument somewhat differently, writing "The evidence indicates that [preventive measures] rarely reduce medical expenditures. Thus prevention is not the answer to rising health care costs. Instead it offers better health at additional cost, squarely posing the question: when is health gain worth the cost?"<sup>2</sup> The goal of this monograph is neither to support nor to refute either of these particular views. Instead, it is to provide objective procedures to assess the aggregate health returns to society of alternative interventions—whether treatment or prevention—thereby allowing such arguments to be judged on their merits as they apply to specific types of health interventions.

### Cost-Utility Analysis

Over the past decade, many researchers have begun to express the outcome of a CEA in terms of a ratio measured in units such as "dollars per quality adjusted life year." This approach, which forges a tight link between CEA and the VonNeumann-Morgenstern (VNM) utility theory, is fast becoming the most widely accepted methodological approach for CEA. This is the method that is sometimes called cost-utility analysis or CUA.

While many researchers and practitioners in the area would be content to consider this *the* legitimate way in which CEA should be conducted, the term "cost-effectiveness" continues to carry multiple meanings. The Comparative Benefits Modeling Project recognized from the outset that it was well ingrained in the general lexicon as a generic and somewhat subjective term. It would be unavailing, not to mention presumptuous, to pronounce that others should henceforth not use CEA in any other way than it is used here or that CEA should be barred from serving objectives other than those addressed here.

---

<sup>1</sup>Norman Daniels, *Just Health Care*, Cambridge, MA:Cambridge University Press, 1985, p. 14.

<sup>2</sup>Introduction to Louise Russell, *Is Prevention Better than Cure*, Washington, DC: Brookings Institution, 1986, p. vii.

This set of difficulties is avoided by usurping the newer term *cost-utility analysis* to describe the form of analysis presented here. This follows the lead of works such as Anderson, et al. (1985), Torrance (1986), and Drummond, et al. (1987), who also use CUA to distinguish this particular form of CEA from others. CUA is developed here as a specific tool of analysis with a distinct framework and orientation. For expositional convenience, this monograph uses the term "cost effective" at times to convey that a given program is judged to be a good use of resources from a CUA perspective; at other times, it uses "cost-utility effective" to help avoid confusion.

The approach to CUA proposed here reflects some fairly strong biases, and these should be recognized from the outset. First, the proposed framework is designed to demonstrate a method that maximizes aggregate social health benefits from a given level of social health care resources. This, in turn, is intended to improve the understanding of decision makers about how the aggregate return to society in health terms of an investment in a given health intervention compares to investment in alternative health interventions.

This goal, admittedly, cannot always or even generally be used to further other legitimate objectives of policymakers. CUA is *not* intended to serve as a general decision aid to further whatever objectives a given decision maker or policy maker happens to have. This is not because such objectives are illegitimate or not proper to take into account in making health-care resource-allocation decisions. It is instead because there is a separate field of analysis (decision analysis) which is designed for this purpose. CUA in this monograph is not considered to subsume decision analysis. On the other hand, it does turn out that the same techniques put forth in the basic model for CUA proposed here can, with minor modifications, sometimes be adopted to promote some of the other, more parochial objectives that policymakers may have.

### Limits of CUA in Health Care

Useful as CUA can be, it is important not to gloss over its limitations. Two deserve special mention. First, CUA by itself cannot resolve important moral and ethical issues in health care. This is not to say that there are not normative and ethical underpinnings to CUA. There are. But, there are important limits as well to what CUA can tell a policy analyst when it comes to moral judgments about health care.

Second, the domain in which CUA can be fruitfully applied should not be exaggerated. In particular, there are many potentially important health interventions that we still do not know enough about; i.e., biological and medical science cannot yet indicate with precision the magnitude of an intervention's likely health outcome. Just because substantial uncertainty exists about the effects of a health intervention does not mean that CUA cannot properly account for this uncertainty; its conclusions may still be valuable. But, the more uncertain the effects of a health intervention the more uncertain will be the conclusion of a CUA. CUA cannot resolve or eradicate uncertainties about basic clinical efficacy. Moreover, there is a danger that the seeming sophistication of a CUA may give an undeserved aura of precision to its conclusions and that caveats about the uncertainty of the conclusions will be lost.

### About This Monograph

These limitations highlight the fact that CUA, like many other tools of analysis, is a field in which a little knowledge is a dangerous thing. In addition to simply presenting a model of CUA in health care, it is important to justify its use and to articulate the strengths and weaknesses of this technique. This monograph, therefore, not only explains how a cost-utility analysis is constructed, but also

explores the assumptions that underlie CUA as applied to health care. It also attempts to provide a clear understanding of what a CUA's conclusions mean and do not mean.

On the other hand, a full presentation of the theoretical underpinnings behind all aspects of CUA would require a volume of a length and technicality beyond that envisioned for this monograph. This volume is somewhat more introductory. In many places the text simply asserts that given properties of the analysis can be shown to hold. It also presumes less specific background on behalf of the reader in economics, multiattribute utility theory, or decision analysis.

This monograph is divided into three parts:

- Part I presents a basic definition of CUA and outlines the steps for applying the mechanics of CUA to evaluating health care interventions.
- Part II steps back from the mechanics of CUA and explores the normative underpinnings of CUA as applied to individual and societal decisions about health care resource allocation. Part II also compares CUA with cost-benefit analysis and decision analysis.
- Part III addresses some of the practical complications that arise in actually performing a CUA.

The framework presented in the chapters that follow is intended to provide a clear and rational perspective on judging competing claims for health care resources. It is hoped that this framework will enhance the usefulness and impact of CUA in the policymaking process.

## **PART I**

### **THE MECHANICS OF CUA IN HEALTH CARE**

Part I defines cost-utility analysis (CUA) as it is applied to health care and outlines methods of using it to make decisions concerning health care interventions. Its first chapter offers an initial definition of CUA. Chapters 2 through 5 focus on the application of CUA in different situations: when deciding between mutually compatible health interventions; when deciding between mutually exclusive interventions; when choosing the extent or magnitude of health interventions; and in other cases, such as those in which the two parts of the cost-effectiveness ratio—the costs and the health outcomes—are not both positive.



## **CHAPTER 1**

### **AN INITIAL DEFINITION OF CUA IN HEALTH CARE**

#### **1. INTRODUCTION: COST-UTILITY ANALYSIS**

This chapter provides an initial, and relatively general, definition of CUA in health care, which will be expanded in subsequent chapters. It also introduces six basic steps of CUA; discusses ways in which CUA is sometimes used (or misused) in the literature; and draws distinctions between CUA and other similar forms of analysis.

CUA is defined in terms of the following three characteristics:

- CUA is a method by which to gauge the desirability of a health care intervention relative to the status quo, using the ratio of the aggregate costs of the intervention to a numerical valuation of the aggregate change in health outcomes that result from it. This ratio is referred to as the "cost-effectiveness ratio."
- Aggregate costs are expressed in monetary units and are the sum of the costs experienced by everyone affected by the health intervention. These costs are measured relative to those associated with the status quo.
- The change in the aggregate health outcomes is the sum of a numerical measure of changes in individual health status for all those whose health is affected by the health intervention.

This definition does not specify how costs should be computed, how health should be measured, what distinguishes a cost from a health effect, or how to incorporate nonhealth effects into the analysis. These are all concerns which will be addressed subsequently.

#### **2. THE BASIC STEPS OF CUA**

During the past dozen years, several authors have proposed steps that should be followed in conducting a cost-effectiveness analysis (CEA) or CUA. Notable among these are Weinstein and Stason (1977), Lave and Lave (1978), Shepard and Thompson (1979), Office of Technology Assessment (1980), Rogers, Eaton, and Bruhn (1981), and Drummond, Stoddart, and Torrance (1987). Their work generally contains the following six basic steps:

Step 1. Define the health intervention under consideration. This involves clearly defining the number and types of the individuals who will experience the intervention, the specific nature of the intervention, and so forth. It also requires a careful presentation of what the health intervention is being compared with. Typically this involves specifying the "status quo" that will prevail if the health intervention is not adopted.

Step 2. Identify the costs of the health intervention. This involves identifying the health care resources and other costs that are associated with the health intervention.

Step 3. Identify the health consequences and other impacts of the health intervention. This involves identifying the impacts of the health intervention on the individuals experiencing them. This also may involve identifying the impacts on other individuals who are indirectly

affected by the health intervention, such as families or employers of the individuals receiving the health care intervention.

Step 4. Measure the costs of the health intervention. This involves applying accepted economic procedures for attaching monetary values to the inputs of the health intervention.

Step 5. Measure the health and other impacts of the health intervention. This involves defining a numerical health scale and using it to measure the health and other outcomes of the health intervention.

Step 6. Account for the uncertainties that may exist in the analysis. This involves sensitivity analysis and other techniques to test the robustness of the results of steps 4 and 5 against the potential uncertainties in the magnitude of the costs and the health impacts of a health intervention.

While the general agreement on these steps may suggest substantial consensus, many of the steps are somewhat vague. In reality, the guides vary somewhat in answering questions such as: What constitutes a cost? What constitutes a health impact? How exactly are costs to be measured? and How exactly are health and other impacts to be measured? Subsequent chapters will explore these areas of controversy.

### **3. NARROWING THE DEFINITION: WHAT "COST-UTILITY EFFECTIVENESS" IS NOT**

It is important also to consider what CUA is not and to differentiate it from other forms of analysis.

#### **3.1. Cost-Utility Effectiveness is Not Synonymous with "Saving Money."**

For some, the term cost effective means that a health intervention saves money without leading to worse health. Such interventions clearly *are* cost effective in the sense that, from a cost-effectiveness perspective, they are preferred to the status quo (not adopting the health intervention). At the same time, it is unduly restrictive to limit CUA to simply determining whether a program saves money without harming health.

A view of cost-utility effectiveness as "saving money without worsening health" rules out interventions that improve health but do not save money or that lower costs but with some additional adverse health impact. Decisions involving these kinds of interventions are the decisions that benefit most from systematic analysis. Restricting the analysis to consider only health interventions that save money without harming health restricts CUA far too narrowly and rules out the very cases where CUA could be most valuable.

Moreover, just because an intervention saves money without worsening health does not mean there are not other competing interventions that are superior to it. Consider two competing (mutually exclusive) programs that, compared to the status quo, save money and improve health. Health Intervention A may save \$500 and improve health by 7 units while Health Intervention B may save \$700 but only improve health by 3 units. While these are the types of problems it is nice to have, it is not obvious which health intervention is preferable. The definition of CUA at the beginning of this chapter allows this issue to be addressed; the narrower definition does not.

### 3.2. CUA Is Not Synonymous with Evaluating the Least Costly Way of Achieving a Given Set of Outcomes

People sometimes view CUA as a method that allows for the comparison of different health interventions, each of which achieves the same objective, in order to decide which intervention accomplishes this objective at the lowest cost. This view, too, is rejected here. If taken literally, this definition reduces the applicability of CUA almost to the vanishing point. In few comparisons of interest do the health interventions that are considered achieve the same objective or result in the same health outcomes.

This is true even in cases in which the health outcomes of different interventions are directly comparable. Consider two programs to encourage individuals to stop smoking. Their objectives are the same—to get smokers to quit smoking. This outcome may be measurable in terms of the proportion of individuals in the program of a given type (e.g., of a given age, sex, or smoking history) that quit smoking for a given time period (say, one year). Even here, however, the view of CUA as a procedure for determining which intervention achieves a given health outcome at lowest cost is suspect. Suppose Program A costs \$100 per participant and results in a 30 percent chance that a given participant will quit smoking while Program B costs \$60 per participant and results in a 10-percent chance that a given participant will quit smoking. As applied to a given population of, say, 1,000 individuals, these two programs do not result in the same health outcomes. Program A would result in 300 individuals stopping smoking while Program B would result in 100 individuals stopping smoking. Requiring CUA to apply only to programs that achieve the same objective would restrict it from being applicable to situations such as this.

More generally, the view that CUA is a methodology to determine which of several interventions achieve a particular health outcome at the lowest cost is also unacceptable because different health interventions typically lead to qualitatively different types of health outcomes. This is true even for interventions designed for the same basic illness. Consider, for example, a cost-effectiveness analysis of different treatments for individuals suffering from chronic depression. Suppose there is a choice between using psychotherapy and drug treatment and that, for argument's sake, these alternatives are mutually exclusive for an individual. Both interventions are designed to treat the same illness. One might imagine that in principle it would be easy to devise a measure of effectiveness to compare the health outcomes of each intervention—e.g., the average number of depressive episodes per year following treatment. Closer inspection indicates that things are not so simple, however. Depressive episodes vary across several salient dimensions, such as initial severity, duration, and level of mental well-being after restabilization. Moreover, drugs may have certain side effects, such as weight gain, lethargy, or sexual dysfunction. Suddenly, we now are in a situation in which probabilities of depression must be traded off against, say, sexual dysfunction or fatigue in the overall assessment of health. Again, we would argue that CUA must be able to address these sorts of situations if it is to be of practical value.

It should be noted, however, that the rejection of CUA as a method designed only to choose the least expensive way to achieve a given health outcome *does* put us at variance with some of the literature. Gramlich, in his cost-benefit textbook, writes:

We can turn the evaluation question around and ask the second type of question: For two alternative programs that accomplish a particular goal, which is cheaper? Here we are performing a comparative, or "cost-effectiveness" evaluation.<sup>3</sup>

Similarly, OTA (1980) writes:

CEA permits comparison of cost per unit of effectiveness among competing program alternatives designed to serve the same basic purpose. Unlike CBA, however, the technique does not allow comparisons of programs having widely different objectives.<sup>4</sup>

This narrower view of CEA also appears in many health economic texts. For instance, Jacobs writes, "Cost-effectiveness studies examine the cost-output relation between alternative methods of achieving the same goal."<sup>5</sup> Similarly, Rapoport, Robertson, and Stuart write that CEA:

...attempts to find which program or approach among several alternatives provides the largest benefit for a given cost or (equivalently) yields a particular benefit (or set of benefits) at the lowest expenditures<sup>6</sup>

### 3.3. CUA is Not an Inventory of Health Outcomes

The foregoing section argues that few health interventions have exactly the same health effects, even when the interventions deal with the same illness. This is even more evident when comparing interventions with different objectives, or with very different approaches to their objectives (e.g., prevention versus treatment). In such circumstances, what can be done when comparing alternative interventions?

One approach is simply to compile the available evidence on the costs of the different interventions and on how such interventions perform on various relevant health outcome dimensions. These then can be presented to a decision maker, who makes whatever comparisons are desired. In this case, no attempt is made to integrate the different dimensions of health outcomes to derive one net "overall health" scale. While such an exercise is not CUA, this does not mean such a listing of health outcomes and costs would not be valuable, even if one went no further. This "inventory" approach, while not providing a particularly compact summary of aggregate costs and health outcomes, undoubtedly would be of use to a policy maker and, in any case, is a necessary starting point for

---

<sup>3</sup>Edward Gramlich, *Benefit-Cost Analysis of Government Programs*, Englewood Cliffs, NJ: Prentice-Hall, 1981, p. 7.

<sup>4</sup>Office of Technology Assessment, op. cit., p. 4. The OTA discussion continues, however, in the following vein: "How are two programs to be compared when one program averts many deaths but has limited effect on disability, and the other prevents considerable disability but averts only a few deaths? Methodological advancements (such as measures of quality adjusted life years) may in time ameliorate this inadequacy, but considerable barriers to using CEA to evaluate programs with significantly different effects still remain."

<sup>5</sup>Philip Jacobs, *The Economics of Health and Medical Care*, Second Edition, Rockville, MD: Aspen, 1987, p. 277.

<sup>6</sup>John Rapoport, Robert Robertson, and Bruce Stuart, *Understanding Health Economics*, Rockville, MD: Aspen, 1982, pp. 495-6.

CUA. Indeed, simply laying out the alternatives facing a policy maker and their characteristics provides many of the advantages of any form of systematic analysis.

The alternative to the "inventory" approach, when health outcomes are not easily compared, is to strive to find a method to measure health which encompasses, on a common, single-dimensional scale, as many health outcomes as possible. Indeed, from the earliest CEAs in health policy, many investigators have felt strongly that if this form of analysis were to provide any overall summary of the effectiveness of different health interventions, some sort of general health outcome scale would have to be constructed. This is implicit, for instance, in Weinstein and Stason's approach (1977), in which CEA is seen as a way to "maximize the total aggregate health benefits conferred" for any given level of resources available. It is also evident in our definition of CUA.

One of the advantages of a general health measure, such as quality-adjusted life-years, mentioned in the introduction, is that such a measure is sufficiently expansive to encompass virtually any health intervention. One can compare such diverse health interventions as prenatal care, treatment for mental illness, and open-heart bypass surgery in terms of cost-utility effectiveness.

Some practitioners and health economists are hesitant to use general measures of health that allow comparisons across a wide range of health interventions. They feel more comfortable restricting CEA/CUA to a comparison of different programs that lead to the same, or comparable, clinical health outcomes and comparing these programs in terms of their respective costs. There is nothing "wrong" or "illegitimate" about this narrower view of the economic evaluation of health interventions. It is important to note, however, that the less willing one is to accept health measurement scales that allow disparate interventions to be compared, the more limited CUA in health care becomes. In the end, without some attempt to make measurement of health outcomes generalizable and comparable, CUA has very little value for policy makers. Conversely, the broader the health measures which can be derived, the greater is CUA's domain and the more useful it is as a guide to resource allocation in health care.

In light of these considerations, CUA is pursued here in the most expansive context in which it can be justified. It is acknowledged that in the end some restrictions may be placed on the breadth of health interventions in which CUA is used. This monograph seeks to identify core conditions under which CUA is applicable and other circumstances for which the proper measurement procedures and interpretations of CUA are likely to be particularly problematic.

Of course, saying that health outcome measures should be able to compare as broad a set of health outcomes as possible does not address the issue of how this can be done. Nor does it indicate the restrictions on such measurements that leave the resulting CUA with normative impact. Clearly, one could establish arbitrarily a numerical, single-dimensional health outcome scale that could, in a completely fabricated way, allow for comparison of different health outcomes. Such a scale would allow for the computation of cost-effectiveness ratios. These ratios, however, would clearly be as worthless as the scale upon which they are based. The challenge, then, is to derive a health measurement scale that encompasses divergent health outcomes and retains normative impact. This challenge is addressed in Part II of this monograph.

### **3.4. CUA is Not DA, COIA, or CBA**

One final aspect of the definition of CUA is its relationship to other forms of analysis, including decision analysis, cost-of-illness analysis, and cost-benefit analysis.

- **Decision analysis (DA)** considers an individual facing a choice, usually one characterized by substantial uncertainties, and uses VonNeumann-Morgenstern utility theory to prescribe the choice that the individual should make.
- **Cost-of-illness analysis (COIA)** is a form of analysis that measures the costs that result from an illness, where the measure of cost is closely tied to national income accounts such as gross national product.
- **Cost-benefit analysis (CBA)** considers a particular program or health intervention and, like CUA, assesses its resource use and health impacts. It is distinguished from CEA and CUA by attaching monetary measures to health outcomes as well as to health care resource use.

Each of these forms of analysis is a distinct methodology, and each has its own strengths and weaknesses. Understanding the relationships between CUA, DA, COIA, and CBA is essential for a full understanding of CUA as it is presented in this monograph. Subsequent chapters describe these relationships in greater depth.

#### 4. CONCLUDING COMMENT

CUA can be seen as an *algorithm*—a method, technique, or set of rules for problem solving. A cost-effectiveness ratio is calculated for a health intervention and its magnitude indicates whether the intervention is cost effective. This definition remains quite general and leaves the following issues to be discussed:

- How can a cost-effectiveness ratio be used to evaluate the desirability of a health intervention relative to the status quo?
- Can CUA be used to make comparisons between different health interventions, rather than only between a health intervention and the status quo?
- Why should a cost-effectiveness ratio for a health intervention be a compelling guide to resource allocation in health care?
- How should costs and health effects be measured in CUA?

These are issues to which we now turn, examining the first two in the remaining chapters of Part I and the latter two in Parts II and III.

## CHAPTER 2

### APPLYING CUA TO EVALUATE MUTUALLY COMPATIBLE HEALTH INTERVENTIONS

#### 1. INTRODUCTION: APPLICATION OF CUA

Chapter 1 defines cost-utility analysis (CUA) as a method to evaluate the desirability of a health intervention relative to the status quo. Turning now to the mechanics—how CUA can be applied—let us first imagine a list of potential health interventions. How can a policy maker decide which ones should be funded and which should not? In other words, how can CUA be used to evaluate the desirability of health interventions relative to the status quo and relative to other health interventions?

To answer this question, we will consider a health care intervention in its simplest and most basic form, as an activity that provides health benefits to some affected group of individuals and consumes health care resources in the process. Also for the purposes of discussion, we will make two assumptions:

- There are no costs other than health care resource costs and no impacts of the health intervention other than the health impacts experienced by the targeted group affected by the health intervention.
- The health interventions under consideration involve positive cost in terms of health care resources and are beneficial to the health of those affected by them—i.e., the numerator and denominator of the cost-effectiveness ratio are both positive.

Both assumptions will be relaxed eventually.

#### 2. MAXIMIZING HEALTH BENEFITS SUBJECT TO A BUDGET CONSTRAINT

##### 2.1. An Algorithm

Notationally, suppose that there are  $K$  different health interventions that might be pursued,  $HI_1, HI_2, \dots, HI_K$ . Each intervention,  $HI_i$ , is associated with a given cost,  $C_i$ , which is measured relative to the costs associated with the status quo. Each intervention leads to an incremental change in aggregate health status  $\Delta H_i$ . Assume for now that the interventions are *mutually compatible*. That is, any subset of interventions is feasible and the costs and health benefits of any of the interventions are independent of which of the other interventions are or are not adopted.

Suppose that there is a total amount of health care resources,  $C_{Tot}$ , to be allocated among these health interventions. This amount  $C_{Tot}$  is sometimes referred to as the *budget constraint*. It again is measured relative to the status quo. Suppose the goal of the resource allocation exercise is *to maximize the total health increment from this given level of resources*. This may well seem to be a difficult task. It can be shown, however, that in general there is a very simple algorithm for reaching this goal.

This algorithm consists of the following steps:

- For each health intervention, form the ratio of its cost to its incremental health benefit,  $R_i = C_i / \Delta H_i$ .

- Rank the health interventions from the lowest to the highest values of  $R_i$ .
- Go down this list and include each health intervention until the budget constraint is just met.

That this algorithm does indeed generally maximize the health benefits from a fixed level of health care resources can be easily demonstrated. Suppose we have ranked health interventions as in step 2 and, without loss of generality, let us denote the first health intervention on this list as  $HI_1$ , the second as  $HI_2$ , and so forth. Suppose that  $HI_5$  has a cost of \$100, an aggregate health benefit of 200, and a cost-effectiveness ratio therefore of 0.5. Suppose that  $HI_7$  has a cost of \$100 also, an aggregate health benefit of 100, and therefore a cost-effectiveness ratio of 1. Consider the situation if  $HI_7$  were funded but not  $HI_5$ . Each dollar spent on  $HI_7$  would yield one unit of incremental aggregate health benefits. If  $HI_7$  were replaced by  $HI_5$ , on the other hand, the same dollars would yield two units of incremental health benefits. Clearly, the aggregate health benefits would be greater from any dollar spent on  $HI_7$  if it were spent on  $HI_5$  instead. Following this logic, it would be desirable in terms of maximizing aggregate health benefits to transfer the \$100 of health care resources from  $HI_7$  to  $HI_5$ .

As this example illustrates, whenever one health intervention has a lower cost-effectiveness ratio than another, dollars spent on the former will return more in aggregate health benefits than dollars spent on the latter. To maximize the aggregate health benefits from a fixed level of funds, one therefore should fund health interventions in reverse order of their ratio of costs to incremental health benefits.

## 2.2. The Link Between the Algorithm and CUA

What does this algorithm have to do with CUA? The two are tightly linked mathematically. To see this, note that when our hypothetical policy maker applies this algorithm, he or she will have one health intervention that is last on the list to be funded. Let  $R(C_{T\alpha})$  be the  $R_i$  value for this last health intervention. We sometimes refer to  $R(C_{T\alpha})$  as the "*threshold cost-effectiveness ratio*." Note also that this threshold ratio is written as a function of  $C_{T\alpha}$ , the total amount of health care resources available. This is because if more health care resources were available, the selection of interventions to fund would extend further down the list and there would be some new, greater, threshold, cost-effectiveness ratio.

Now consider some new or previously unconsidered health intervention. Should it be adopted? That is, would funding it be preferable to the status quo of not having it adopted? To answer this question one could go back and redo the list, with this new intervention included, and see if it gets adopted. But this is not necessary. If the cost-effectiveness ratio of the new health intervention is greater than  $R(C_{T\alpha})$  then in this new listing it would not get funded. The total funds available would be depleted before the policy maker got down the list to the new intervention. On the other hand, if the cost-effectiveness ratio of the new intervention is lower than  $R(C_{T\alpha})$ , the new intervention would be funded. Because previously the algorithm made it down to all programs with cost-effectiveness ratios lower than  $R(C_{T\alpha})$ , the new health intervention would be encountered before the total of the available funds were depleted.

Thus, if one wanted to maximize the health benefits from a fixed level of funds, one could adopt this algorithm and evaluate a health intervention relative to the status quo based purely on its cost-effectiveness ratio. If its cost-effectiveness ratio is lower than  $R(C_{T\alpha})$  then it can be judged cost effective relative to the status quo while if its cost-effectiveness ratio is greater than  $R(C_{T\alpha})$  it can be judged not to be cost effective relative to the status quo. But this is exactly the procedure implied by CUA, as specified in our basic definition of CUA. We can also reverse this chain of logic. If a

health intervention can be compared to the status quo based on its cost-effectiveness ratio then there must be some threshold ratio that distinguishes cost-effective from cost-ineffective health interventions. This means that, for any set of mutually compatible health interventions, all health interventions with a cost-effectiveness ratio below this threshold are adopted. This is the same procedure that maximizes the aggregate incremental health benefits from a given level of funding.

Thus, defining CUA as a method by which a health intervention can be judged relative to the status quo based on its cost-effectiveness ratio implies that CUA serves to allocate health care resources to maximize the aggregate health care benefit. Put differently, regardless of whether a policy maker consciously has this objective in mind, if health interventions are evaluated relative to the status quo based on their cost-effectiveness ratios, this is consistent with maximizing health benefits subject to a budget constraint.

Note that this analysis does not indicate the proper level of total spending,  $C_{T\alpha}$ , that should be devoted to the  $K$  health interventions. This will be discussed in subsequent chapters. However, whatever level of  $C_{T\alpha}$  is used, the ordering of the health interventions from best to worst remains the same. The only question is how far "down the list" one should go. The larger the budget, the further down the list one should go in funding programs and the greater the threshold cost-effectiveness ratio.

### 2.3. An Illustration

These points can be illustrated by the following numerical example. Suppose there are the following six governmental health interventions (Table 1), with costs and changes in health status (both measured relative to the status quo). Subject to its budget constraint, the Government can fund any subset of them.

**Table 1. A Set of Mutually Compatible Health Interventions**

| Health Intervention | Cost | Change in Health Status |
|---------------------|------|-------------------------|
| 1                   | \$30 | 30                      |
| 2                   | 30   | 20                      |
| 3                   | 55   | 70                      |
| 4                   | 20   | 40                      |
| 5                   | 30   | 50                      |
| 6                   | 30   | 75                      |

To determine which health interventions to fund, the first step is to reorder these interventions in terms of their cost-effectiveness ratios (Table 2).

**Table 2. Health Interventions Reordered by Cost-Effectiveness Ratios**

| Health Intervention | Cost | Change in Health Status | Cost-effectiveness Ratios |
|---------------------|------|-------------------------|---------------------------|
| 4                   | \$20 | 40                      | .5                        |
| 5                   | 30   | 50                      | .6                        |
| 6                   | 50   | 75                      | .666                      |
| 3                   | 55   | 70                      | .786                      |
| 1                   | 30   | 30                      | 1.0                       |
| 2                   | 30   | 20                      | 1.5                       |

Suppose there is \$100 to be spent on some subset of the six health interventions. Then following our algorithm, Health Interventions 4, 5, and 6—the three interventions with the lowest ratios of cost to change in health outcomes—should be chosen. Funding this set of programs will cost a total of \$100 and will provide 165 units of incremental health improvement. No other set of interventions will require \$100 or fewer units of resources and result in as much aggregate improvement in health.

If there is a new health intervention to consider, one need only compare its cost-effectiveness ratio with the existing threshold ratio to determine if the program should be adopted—i.e., if it is preferred to the status quo. Suppose there is \$100 to spend and a new health care option, HI<sub>7</sub>, arises. It costs, say \$50, and provides 80 units of aggregate health benefits. If we were to add it to the above list, then it should be funded, along with HI<sub>4</sub> and HI<sub>5</sub>. We need not actually go to the effort of putting it on the list and working through our algorithm. We know it will be cost effective simply because its cost-effectiveness ratio of .625 is below the threshold ratio of .666. Similarly, if HI<sub>7</sub> had health benefits of only 70 we would know that it would not be cost effective since now its cost-effectiveness ratio of .71 is above the threshold ratio.

If the total budget for the six health interventions changes, the procedure to rechoose which health interventions to fund is straightforward. For example, if there is \$155 to be spent, we follow the same algorithm but now we can include Health Intervention 3 along with 4, 5, and 6 as cost-effective. As indicated above, as the total amount to be spent changes, all that happens is that the decision maker goes further down the list of health interventions in choosing which ones to fund. As can also be easily seen, when there is more to be spent, the threshold cost effectiveness rises. In this example, it rises from .666 when there is \$100 to spend to .786 when there is \$155 to spend.

## 2.4. Minimizing Costs Subject to an Aggregate Health Improvement Constraint

It is useful to note that the same rules and procedures apply if CUA is viewed as a method to minimize the required health care resources for a given level of total health benefits. For instance, in the numerical example given above, suppose the objective were to minimize the cost of providing 165 units of health benefits. The solution to this objective would be to fund Health Interventions 4, 5, and 6, leading to a cost of \$100. No other set of health interventions provides this level of health benefits for any lower cost. Similarly, if the objective were to minimize the cost of providing 235 units of health benefits, the solution to this objective would be to fund Health Interventions 4, 5, 6, and 3. No other set of health interventions provides this level of benefits for any lower cost. These solutions, however, are exactly the same ones we reached when trying to maximize health benefits subject to a budget constraint.

Cost-effectiveness analysis of health interventions can be justified either as maximizing health benefits subject to a budget constraint or minimizing costs subject to a health benefit constraint. Neither is "better" or "worse" than the other as a way to frame cost-effectiveness analysis. Mathematically they are completely equivalent formulations. However, we find it more natural to think in terms of maximizing health benefits, particularly in situations in which health care resources may be limited by political or bureaucratic considerations, and it is appropriate for a Government bureau or agency to try to maximize the health benefits these limited resources can provide. In contrast, there are fewer instances in which a political or other external constraint imposes a set level of health benefits and directs a Government office to minimize the costs of achieving them. In any case, throughout this and subsequent chapters, we will follow the norm of considering how to maximize health benefits subject to a budget constraint.

## 2.5. A Complication: The "Lumpiness" of Potential Interventions

It was indicated above that "in general" the way to maximize aggregate health benefits subject to a budget constraint is to rank health interventions in order from the lowest to the highest in terms of the ratio of cost to change in health outcome. There are some instances when this will not be true.

For example, suppose that the budget constraint is \$80. Now, going down the list as before, we choose Health Interventions 4 and 5, providing 90 units of health improvement and costing \$50. But choosing the next health intervention, Health Intervention 6, would cause us to exceed the \$80 budget constraint. Consider, though, skipping Health Intervention 6 and Health Intervention 3, which also costs too much, and going to the first intervention that does fit within the budget constraint, Health Intervention 1. If the Government funded Health Interventions 4, 5, and 1 the cost would be \$80 and 120 units of health improvement would result.

This may seem to be the procedure that should be followed. If we are to maximize the health benefits from the \$80 however, this would be the wrong set of health interventions to finance. Instead, Health Interventions 5 and 6 should be selected, providing 125 units of health improvement from the \$80.

Why is it that the algorithm, based solely on cost-effectiveness ratios, does not succeed in maximizing health benefits subject to a budget constraint when the budget constraint happens to be \$80, instead of \$100 and \$155? What is wrong with the \$80 budget constraint? The difficulty is caused by a form of "lumpiness." Moving from Health Interventions 4 and 5, costing \$50 together, there was \$30 left within the budget constraint. But the next health intervention, Health Intervention 6, was too "lumpy" to fit within the remaining funds.

In the presence of this kind of "lumpiness," we might need numerical optimization methods to determine the set of health interventions that maximizes health benefits available within a limited budget. Fortunately, this kind of lumpiness is the exception in most cost-effectiveness contexts, and the original algorithm will generally be quite sufficient for determining which health interventions are cost effective and which are not.

There are two reasons for this. First, in many situations the overall amount to be spent on a set of health interventions is very large compared to the cost of any individual component. When this is the case, the type of lumpiness encountered in the example can be shown to be less a concern and will have only a minor impact in terms of modifying the algorithm. In particular, it will modify the algorithm only for health interventions that have cost-effectiveness ratios very near the threshold cost-effectiveness ratio.

Second, health interventions have been portrayed as discrete, finite alternatives, but in reality, health interventions of a given type can vary in terms of their size or scale. If there is only \$80 to be spent in total, and Health Interventions 4 and 5 consume \$50, Health Intervention 6 cannot be made to fit within the budget limit. But, typically Health Intervention 6 could be scaled down. Perhaps there is a variant of Health Intervention 6 that costs only \$30 (and provides some level of health benefits now less than 75). When reduced or expanded variants of health interventions are possible and when such trimmings and expansions alter the associated level of health benefits in a regular and "smooth" fashion, then the lumpiness problem encountered generally will cease to be a problem.

## CHAPTER 3

### APPLYING CUA TO EVALUATE MUTUALLY EXCLUSIVE HEALTH INTERVENTIONS

#### 1. INTRODUCTION

In Chapter 2 we assumed that any subset of health interventions could be pursued simultaneously. What happens, however, if this assumption is not valid? In this chapter, we examine the basic mechanics of cost-utility analysis (CUA) when the health interventions under consideration are mutually exclusive.

This situation is quite common. When someone is suffering from a particular illness they may be treated either by drugs or by surgical intervention. It may not make sense to do both. For example, treatment of gallstones may be by pharmacologic therapy and lithotripsy in order to dissolve and break down gallstones. Alternatively, it may involve surgical removal of the gall bladder. It would be redundant to treat a given individual with both modes of therapy.

#### 2. A BASIC ALGORITHM

With mutually exclusive health interventions, the objective of the overall analysis remains the same even if the mechanics are a bit modified. The goal remains to provide the most health benefits from a given level of resources.

We present two algorithms for determining which among a set of mutually exclusive health interventions is most cost effective. Each leads to the same result, although the algorithms differ in form, each shedding its own light on the logic behind dealing with mutually exclusive interventions.

The first algorithm for dealing with mutually exclusive health interventions is the following.<sup>7</sup>

1. Rank the health interventions in terms of their cost, going from lowest to highest cost intervention. Include among the list the status quo intervention  $HI_{sq}$ , which has zero cost and zero incremental health benefits. If two interventions have the same cost, list the one with lower health benefits first.
2. Remove from the list any health intervention that costs the same or more than the prior intervention on the list but produces lower incremental health benefits.
3. Beginning with the second intervention on the list, compare each intervention with the prior one on the list. Compute the difference in cost between the two interventions and the difference in incremental health benefits. Call the ratio of these differences the *incremental cost-effectiveness (CE) ratio*.

For instance, assume, without loss of generality, that after step 2 the Health Interventions are numbered sequentially and suppose Health Intervention 4 costs  $C_4$  and produces health benefits  $\Delta H_4$ , and that Health Intervention 3

---

<sup>7</sup>Appreciation is expressed to Milton Weinstein for discussions which led to the refinement and improvement of this algorithm compared to an earlier version.

costs  $C_3$  and produces health benefits  $\Delta H_3$ . Then the incremental cost-effectiveness ratio for Health Intervention 4 is:  $(C_4 - C_3)/(\Delta H_4 - \Delta H_3)$ .

4. Eliminate from the list any health intervention with an incremental cost-effectiveness ratio greater than that of the next health intervention on the list.
5. Go down the remaining list until the last health intervention with an incremental cost-effectiveness ratio less than the threshold ratio,  $R_{\text{Threshold}}$ , is reached. This is the desired health intervention in terms of cost effectiveness.

This algorithm for mutually exclusive health interventions may seem a bit convoluted. In fact, it is fairly straightforward. The first four steps essentially prune from the set of possible health interventions those that are *dominated* by some other intervention. In this context, we can define two forms of cost-effective dominance: *strong cost-effective dominance* and *weak cost-effective dominance*. We discuss these concepts briefly.

In strong cost-effective dominance one health intervention is superior to another on both the cost and health outcome dimensions. For instance, Health Intervention A may cost less and result in larger incremental health benefits than Health Intervention B. While the world seldom works so neatly that all comparisons among health interventions are of this sort, when situations such as this arise, Health Intervention A clearly is preferred to Health Intervention B in terms of cost-effectiveness. In this case, Health Intervention A is said to (strongly) dominate Health Intervention B from a cost-effectiveness perspective. Step 2 in this algorithm eliminates from the list of candidate health interventions those that are strongly dominated by some other health intervention.

After eliminating the strongly dominated health interventions, the list consists of interventions that sequentially cost additional money and provide additional health benefits. The question becomes whether the additional health benefits are worth the additional cost. To answer this, incremental cost-effectiveness ratios are calculated, indicating the additional costs per additional unit of health benefit as one moves from one health intervention on the list to the next.

Even having computed the incremental cost-effectiveness ratios, it may appear that it is not possible to prune the list further without knowledge of the threshold cost-effectiveness ratio,  $R_{\text{Threshold}}$ . In fact, though, additional pruning is often possible. To see this, suppose the incremental cost-effectiveness ratio of Health Intervention 2 is 1.0 while the incremental cost-effectiveness ratio of Health Intervention 3, the next one on the list, is 0.6.

Suppose moving from Health Intervention 1 to Health Intervention 2 is judged cost effective. That is, the incremental cost-effectiveness ratio of Health Intervention 2 is below the threshold cost-effectiveness ratio. If this is the case, then the incremental cost-effectiveness ratio of Health Intervention 3, being below that of Health Intervention 2, will also be below the threshold cost-effectiveness ratio. As such, the move from Health Intervention 2 to Health Intervention 3 will also be cost effective. Thus, Health Intervention 2 can be rejected in favor of Health Intervention 3.

Generalizing this example, whenever the incremental cost-effectiveness ratio of Health Intervention  $i$  is above that of Health Intervention  $i+1$ , we can eliminate Health Intervention  $i$  from our list of potential health interventions. Regardless of the threshold cost-effectiveness ratio, if it is cost-effective to do Health Intervention  $i$  relative to Health Intervention  $i-1$ , it will be cost-effective to do Health Intervention  $i+1$  instead.

This is what is meant by weak cost-effective dominance. Health Intervention A weakly dominates Health Intervention B from a cost-effectiveness perspective if Health Intervention A will be preferred to Health Intervention B regardless of the threshold cost-effectiveness ratio. Step 4 prunes the list of potential health interventions by eliminating those that are (weakly) dominated by other interventions.

Finally, having pruned from the list of potential health interventions those that are strongly and weakly dominated, we can proceed no further without resorting to a particular value for the threshold cost-effectiveness ratio. Having picked a numerical value for  $R_{\text{Threshold}}$ , one must go down the list of remaining health interventions until the last intervention is reached with an incremental cost-effectiveness ratio less than  $R_{\text{Threshold}}$ . Up to this point, as one goes down the list of health interventions, the incremental health benefits have been worth the incremental costs. After this point, this is no longer so.

Note that this algorithm does not determine the appropriate value of  $R_{\text{Threshold}}$ . It takes this value as given, just as the total amount of health care resources available was taken as given in the prior chapter. Indeed, we imagine the choice among mutually exclusive interventions to be taking place in a setting where health care resources also are being allocated among mutually compatible programs. In both cases,  $R_{\text{Threshold}}$  typically would be determined by the same type of procedure. It therefore would depend ultimately on the level of available resources.

### 3. A NUMERICAL EXAMPLE

This algorithm can be illustrated by the numerical example of Section 2, now assuming that Health Interventions 1 through 6 are mutually exclusive and only one of them can be chosen. We begin by ordering the interventions in terms of cost (Table 3).

**Table 3. Mutually Exclusive Health Interventions**

| Health Intervention | Cost | Change in Health Status |
|---------------------|------|-------------------------|
| 0 (status quo)      | \$ 0 | 0                       |
| 4                   | 20   | 40                      |
| 2                   | 30   | 20                      |
| 1                   | 30   | 30                      |
| 5                   | 30   | 50                      |
| 6                   | 50   | 75                      |
| 3                   | 55   | 70                      |

Health Intervention 2 is strongly dominated by Health Intervention 4, costing more money and providing fewer health benefits. Eliminating Health Intervention 2 from the list, Health Intervention 1 is also strongly dominated by Health Intervention 4. Finally, Health Intervention 3 is strongly dominated by Health Intervention 6. Removing Health Interventions 2, 1, and 3 from the list yields the following table with the incremental cost-effectiveness ratios calculated for the remaining interventions.

**Table 4. Health Interventions Pruned of Strongly Dominated Interventions**

| Health Intervention | Cost | Change in Health Status | Incremental CE Ratio |
|---------------------|------|-------------------------|----------------------|
| 0 (status quo)      | \$ 0 | 0                       |                      |
| 4                   | 20   | 40                      | .5                   |
| 5                   | 30   | 50                      | 1.0                  |
| 6                   | 50   | 75                      | .8                   |

Applying step 4, Health Intervention 5 is weakly dominated by Health Intervention 6. No matter what the threshold cost-effectiveness ratio, if it is worth moving down the list from Health Intervention 4 to Health Intervention 5, it will also be worth moving down the list from Health Intervention 5 to Health Intervention 6. Health Intervention 5 can therefore be removed from the list, yielding the following table.

**Table 5. Health Interventions Pruned of Strongly and Weakly Dominated Interventions**

| Health Intervention | Cost | Change in Health Status | Incremental CE Ratio |
|---------------------|------|-------------------------|----------------------|
| 0 (status quo)      | \$ 0 | 0                       |                      |
| 4                   | 20   | 40                      | .5                   |
| 6                   | 50   | 75                      | .857                 |

Having applied steps 1 through 4 of the algorithm, the selection of the cost-effective intervention now requires specification of the threshold cost-effectiveness ratio. Notice, however, that the list of potential interventions has been refined considerably without needing to specify  $R_{\text{Threshold}}$ .

If  $R_{\text{Threshold}}$  is below .5, then none of the above health interventions is cost effective and from a cost-effectiveness perspective the status quo, with its zero level of cost and zero level of incremental health benefits, is preferred. For none of the health interventions is the incremental health benefit worth the cost. If  $R_{\text{Threshold}}$  is between .5 and .857 then Health Intervention 4 should be selected from a cost-effectiveness perspective. Its costs are worth the health benefits it confers relative to the status quo. On the other hand, it is not worth paying the extra amount required by Health Intervention 6. Finally, if  $R_{\text{Threshold}}$  is greater than .857, then Health Intervention 6 should be selected.

The fact that the preferred health intervention among a set of mutually exclusive options may, as in this example, depend on the threshold cost-effectiveness ratio is not surprising. A move from a less costly to a more costly health intervention will be cost effective only if the cost-effectiveness threshold ratio is high enough that the additional health benefits are worth the additional cost.

It may seem that the choice among mutually exclusive health interventions is not only complicated but has strayed significantly from our original definition of CUA. In fact, however, our original definition remains valid. That definition defined CUA as a methodology in which a health intervention is compared with the status quo based on the ratio of the cost of the intervention to changes in the aggregate health status of those affected by the intervention. In dealing with mutually exclusive health interventions that definition remains completely valid. If a given health intervention among the list of mutually exclusive ones is to be compared with the status quo to determine if it is cost effective, then all that is required is information about that intervention's cost-effectiveness ratio. The complications introduced in this chapter arise only when one wishes to move beyond a comparison of an intervention with the status quo to a comparison among different, mutually exclusive interventions.

#### 4. AN ALTERNATIVE SET OF ALGORITHMS

The algorithm provided above can be a bit cumbersome. There is an alternative algorithm that leads to exactly the same choices in all situations but which may be more straightforward to implement as long as one is willing to specify from the outset a particular value for the threshold cost-effectiveness ratio. This algorithm is as follows:

- For each of the mutually exclusive health interventions,  $HI_i$ , compute  $D_i = (\Delta H_i \times R_{\text{Threshold}}) - C_i$ . " $D_i$ " can be thought of as the *difference* between the "imputed" value of the health intervention and its cost. To see this, note that the first of the two terms on the right hand side of the equation that defines  $D_i$  is the imputed value of the health intervention. This is computed as the number of improved health units resulting from the health intervention multiplied by the threshold value associated with each such unit. The second of the terms on the right hand side of the equation that defines  $D_i$  is the cost of the intervention. So  $D_i$  is the difference between these two terms and thus represents the net benefit of the health intervention. Note that the status quo, with a zero level for costs and health benefits, has a  $D_i$  of 0.
- Choose that health intervention with the largest value of  $D_i$ .

A formal proof that this algorithm gives the same results as the more complicated algorithm is not provided here. This can be seen, however, through illustration. Using the same numerical example as above, assume again that Health Interventions 1 through 6 are mutually exclusive and that only one of them can be chosen. For purposes of illustration, we assume that  $R_{\text{Threshold}}$  is .666 (Table 6).

**Table 6.** Choosing From Among Mutually Incompatible Interventions Using the Second of the Algorithms, Assuming  $R_{\text{Threshold}} = .666$

| Health Intervention | Cost | Change in Health Status | $(\Delta H_i \times R) - C_i$ |
|---------------------|------|-------------------------|-------------------------------|
| 0 (status quo)      | \$ 0 | 0                       | 0                             |
| 1                   | 30   | 30                      | -10                           |
| 2                   | 30   | 20                      | -16.666                       |
| 3                   | 55   | 70                      | - 8.333                       |
| 4                   | 20   | 40                      | 6.666                         |
| 5                   | 30   | 50                      | 3.333                         |
| 6                   | 50   | 75                      | 0                             |

Based on the second algorithm, the preferred health intervention is Health Intervention 4 which has a greater value of  $D_i$  (6.666) than any other of the interventions. This is the same health intervention recommended using the prior algorithm for a threshold cost-effectiveness ratio of .666.

Not only can this algorithm be used to address the choice among mutually exclusive health interventions, but it can be modified to choose among mutually compatible health interventions as considered in Chapter 2. For that situation we can use the following algorithm:

- Compute  $D_i = (\Delta H_i \times R_{\text{Threshold}}) - C_i$ .
- Fund any intervention for which the resulting  $D$  value is positive.

While again, a formal proof of this equivalence will not be provided, an illustration is easily given. Assume that the six health interventions considered above are mutually compatible. Under the new algorithm, three health interventions should be funded—Health Interventions 4, 5, and 6, because they are the ones with nonnegative  $D_i$  values. But these are exactly the same health interventions that were recommended in Chapter 2 for the example in which there was a total budget constraint of \$100, a constraint that implies a threshold cost-effectiveness ratio of .666.

Because this second algorithm appears simpler when applied to either mutually exclusive or mutually compatible health interventions, one may inquire why the first algorithm should be used at all. The main reason is that the first algorithm pushes the analysis as far as possible without introducing a specific numerical value for the threshold cost-effectiveness ratio. Often this threshold value is subject to differing views. It is best in such cases not to embed a given value for  $R_{\text{Threshold}}$  deep within the arithmetic of the analysis but rather to leave it for the end where its implications for the conclusions of a cost-effectiveness analysis can be seen. For this reason, we generally prefer the first algorithm to the second. If one is in a situation, however, in which the threshold cost-effectiveness ratio is well established and not subject to controversy, then the second algorithm can lead to the same answer more quickly and neatly.

Note also that the second algorithm shares at least a surface similarity with cost-benefit analysis (CBA). Without going into a lengthy presentation of CBA, we can say that one might imagine health being valued in dollar terms at  $R_{\text{Threshold}}$  per unit.  $D_i$  then takes on the appearance of "net monetary benefits," the key evaluative criterion in CBA. There is something to this interpretation, but it can also be a bit misleading as will be seen in subsequent chapters.



## CHAPTER 4

### APPLYING CUA TO CHOOSING THE MAGNITUDE OF HEALTH INTERVENTIONS

#### 1. INTRODUCTION

Many types of health interventions can be expanded or contracted in scale, as mentioned in the discussion of "lumpy" interventions in Chapter 2. One way a health intervention can be expanded in scale is by applying a given treatment or prevention to a larger pool of individuals. Another way a health intervention can be altered in scale is by varying the intensity of the treatment per individual. For instance, the number of visits to a psychologist per week can be altered. Here the issue is not only whether a given intervention should be pursued, but also to what degree it should be pursued. That is, the question is not only "whether" but also "how much."

#### 2. THE CONDITION $MC/MB = R_{\text{THRESHOLD}}$

Figures 4-1a, 4-1b, and 4-1c illustrate this type of situation. In each figure, the horizontal axis is the scale of some given health intervention. This scale is measured for now in terms of the cost of the intervention. Figure 4-1a shows the total health benefits associated with different scales of the intervention. The total health benefits (TB) are the increment in aggregate health that results from the intervention at each given size. Figure 4-1b shows the associated marginal health benefit curve. Loosely speaking, marginal health benefits can be defined as the additional level of health benefits that result from an additional \$1 of expenditure. More technically, denoting the level of total health benefits associated with scale or cost  $C$  by  $TB(C)$  [where " $TB(C)$ " is the mathematical notation for the total benefits associated with a cost or scale size  $C$ ], the marginal health benefit (MB) associated with cost or scale size  $C$  can be defined as:

$$MB(C) = dTB(C)/dC.$$

For those more graphically inclined,  $MB(C)$  can also be defined as the slope of the total benefit curve in Figure 4-1a at total level of expenditure  $C$ .

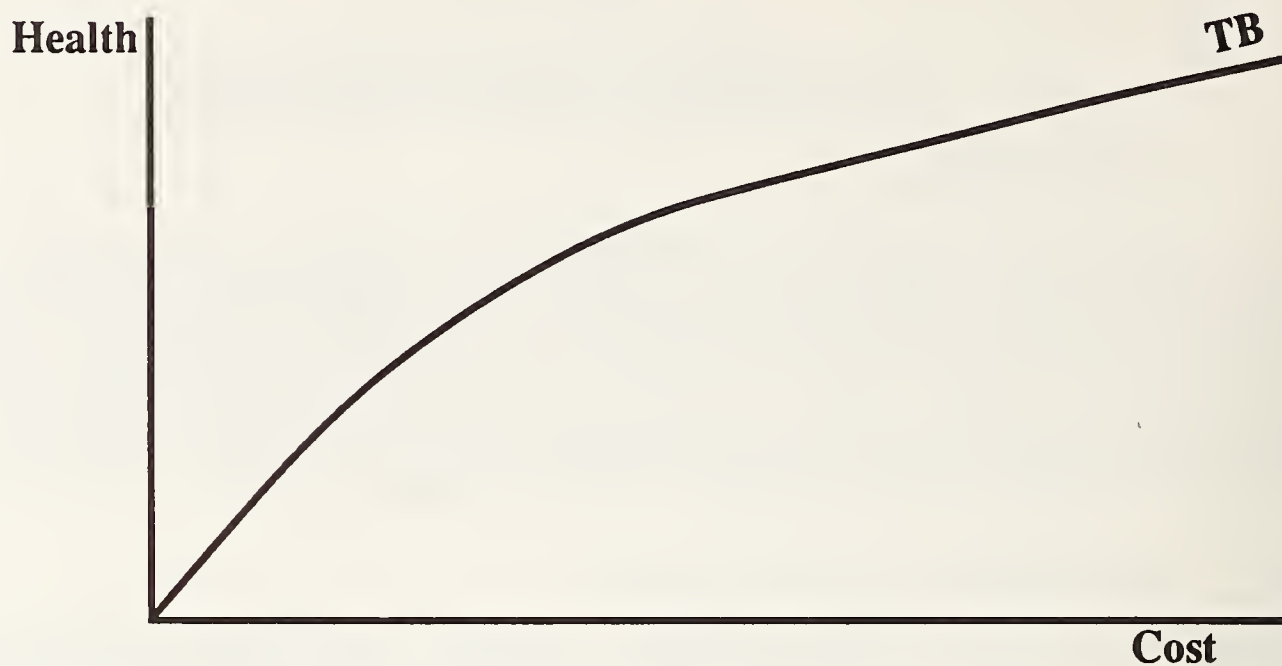
As drawn here, the TB curve is concave and the MB curve decreases as the scale of the intervention increases. This would be the case if, for instance, the scale or cost of the health intervention is increased by including new individuals for whom the health intervention is likely to be less effective. Or, if the scale refers to the intensity of treatment for a given individual, this would be the case if, say, increasing the number of visits per week to a therapist continues to improve health, but the incremental benefit per visit goes down the more visits per week the patient has.

Analogous to an overall cost-effectiveness ratio and the incremental cost-effectiveness ratio, a *marginal cost-effectiveness ratio* can be defined. It can be defined as  $MC/MB$  and is drawn in Figure 4-1c. Note that in this case, where scale is defined in terms of costs, the marginal cost-effectiveness ratio is just  $1/MB$ , the inverse of MB. Figure 4-1c also indicates the average or overall cost-effectiveness ratio for the intervention at each scale, namely  $TC/TB(C)$ .

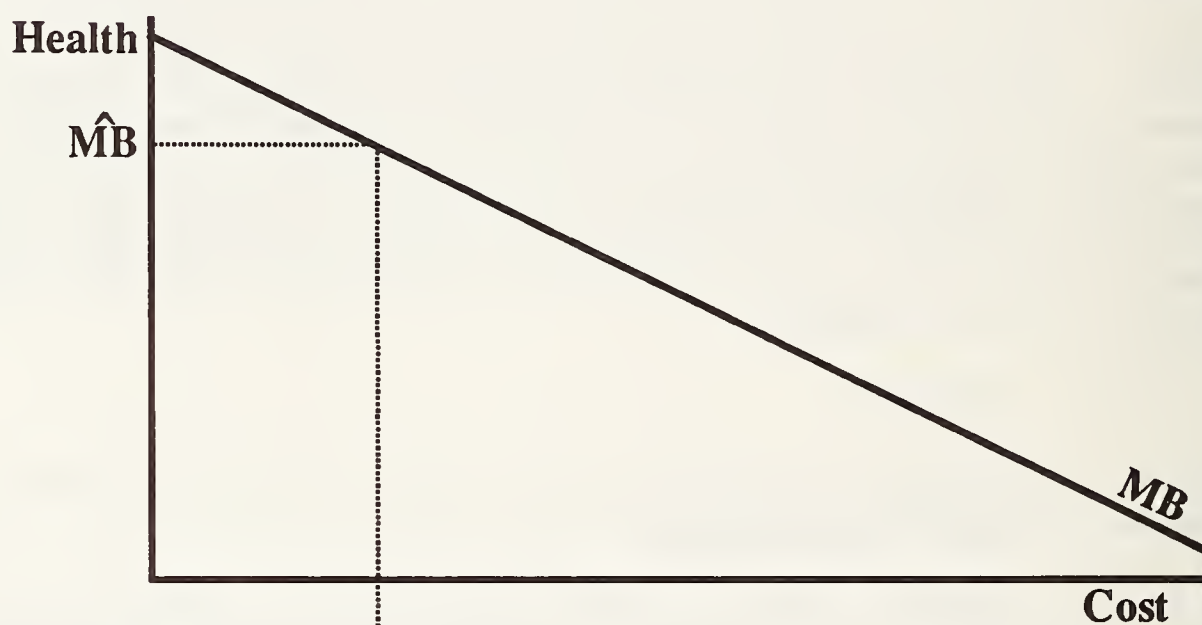
In a case such as that depicted in Figure 4-1, the level of Health Intervention A that is preferred from a cost-effectiveness perspective is the level at which the marginal health benefit from the last unit of expenditure is equal to the threshold level,  $R_{\text{Threshold}}$ . If the threshold level is, say,  $R'$ , as in Figure 4-1c, then the cost-effective scale of the intervention is level A'. To see this, consider some level

Figures 4-1a, 4-1b, 4-1c. Costs, Health Benefits, and Cost-Effectiveness Ratios for Hypothetical Health Intervention A

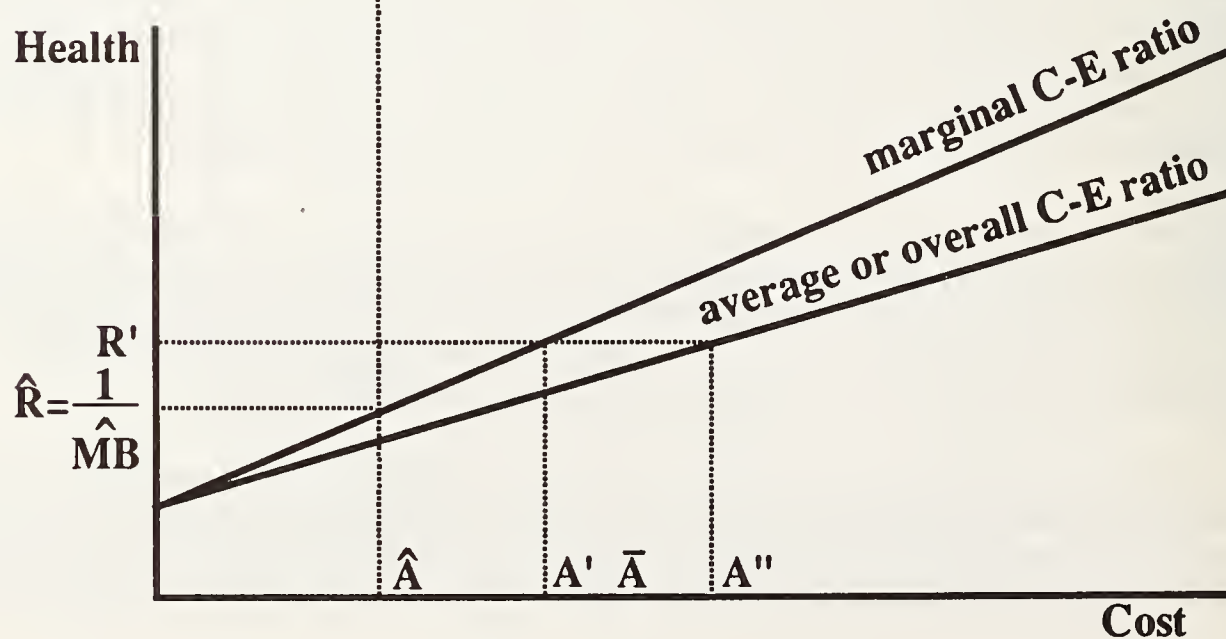
4-1a



4-1b



4-1c



short of  $A'$ , such as  $\hat{A}$ . At  $\hat{A}$ , if one more dollar is spent on the intervention, there will be incremental health benefits, as read off the marginal health benefits curve, of  $\hat{MB}$ . The cost-effectiveness ratio of this additional dollar of expenditure will be  $\hat{R}$ . Since  $\hat{R}$  is less than  $R_{\text{Threshold}}$ , this additional dollar of expenditure is cost-effective. Sufficient incremental health benefits are reaped to justify the additional expenditure. On the other hand, using similar reasoning, for any expenditure past  $A'$ , the additional health benefits are not sufficient to justify the additional cost.

Thus, the cost-effective scale of the intervention occurs when  $MC/MB = R_{\text{Threshold}}$ . With  $MC = 1$ , this condition becomes just  $1/MB = R_{\text{Threshold}}$ . Indeed, this is a general necessary condition for the scale of an intervention to be cost effective. This is easy to see. If  $1/MB < R_{\text{Threshold}}$ , then the health benefits to an additional dollar of spending is sufficient to make that expenditure cost effective. As such, it is not optimal to be in a situation where  $1/MB < R_{\text{Threshold}}$ . Instead an additional amount should be spent on the intervention. Conversely, if  $1/MB > R_{\text{Threshold}}$ , then the health benefits from the last dollar spent are not sufficient to justify it, and the intervention is too large.

As this discussion indicates, in determining the proper scale of an intervention, the focus shifts from looking at the overall cost-effectiveness ratio of an intervention at a given scale and to looking instead at the marginal cost-effectiveness ratio. Note that the need to move beyond overall cost-effectiveness ratios to consider marginal cost-effectiveness ratios only arises in choosing at which level a health intervention should be funded. If we wished to address the more limited question of whether a health intervention at a given level of funding should be pursued compared to the status quo of doing nothing, then all that is required is the computation of the overall cost-effectiveness ratio at that particular level of funding. As shown in Figure 4-1c, for any level of funding less than  $A''$ , the basic cost-effectiveness ratio of the health intervention is less than the threshold level. If faced with a discrete decision to fund or not fund, any level of funding below  $A''$  is cost effective.

A curious feature of the two curves in Figure 4-1c is that  $A''$  is greater than  $A'$ . This means that there are levels of scale for Health Intervention A which are greater than the most desirable scale of funding,  $A'$ , but which are still cost-effective compared to the status quo. Level of funding  $\bar{A}$  is an example of such a level. Upon reflection, this property is not surprising. Size  $\bar{A}$  is larger than the most cost-effective level. The additional expenditures in moving from size  $A$  to  $\bar{A}$  do not provide enough additional benefits to justify them from a cost-effectiveness perspective. Nonetheless, if the only alternative to level  $\bar{A}$  is to do nothing at all and leave things to the status quo, then level  $\bar{A}$  is cost effective. It provides enough additional benefits compared to doing nothing (as opposed to being compared to expenditure level  $A'$ ) to justify the expenditure.

### 3. $MC/MB = R_{\text{THRESHOLD}}$ IS A NECESSARY BUT NOT SUFFICIENT CONDITION

As shown in Figure 4-1c, the marginal cost-effectiveness ratio curve lies above the average cost-effectiveness ratio curve and is positively sloped throughout. This is not always the case. For instance, it may be that a diagnostic test uses a very expensive piece of equipment. It costs a lot per patient to process only a few people since the entire piece of equipment must be purchased. Servicing many more patients may cost very little additional in total and lead to a much lower average cost. Each additional patient may cost the same as the previous patient. As more and more patients are seen, the patients have less to gain from the diagnostic test. As such, the marginal benefit curve slowly declines.

The following chart provides a numerical example of this situation for a health intervention denoted as Health Intervention B. The initial patient costs \$100, while additional patients cost only \$2 each. The incremental health benefit slowly declines from 20 units for the first patient down to 6 units per

patient by the time the 202<sup>nd</sup> patient is served. As a result, the marginal cost-effectiveness ratio slowly rises from .105 for the second patient to .333 for the 202<sup>nd</sup> patient. Because of the initial expense of treating the first patient, the overall cost-effectiveness ratio begins quite high, at 5, declines through the 101<sup>st</sup> patient, and then slowly rises. The overall cost-effectiveness ratio remains above the marginal cost-effectiveness ratio through the 101<sup>st</sup> patient (Table 7).

**Table 7. Cost-Effectiveness Ratios for Health Intervention B**

| No. of Patients | Cost  | Total Health Benefits | Overall C-E Ratio | Marginal C-E Ratio             |
|-----------------|-------|-----------------------|-------------------|--------------------------------|
| 1               | \$100 | 30                    | 5                 | $100/20 = 5.000$               |
| 2               | \$102 | 30                    | 2.61              | $(102-100)/(39-20) = .105$     |
| 3               | \$104 | 57                    | 1.82              | $(104-102)/(57-39) = .111$     |
| ...             |       |                       |                   |                                |
| 50              | \$198 | 792                   | .25               |                                |
| 51              | \$200 | 807                   | .248              | $(200-198)/(807-792) = .133$   |
| 52              | \$202 | 821                   | .246              | $(202-200)/(821-807) = .143$   |
| ...             |       |                       |                   |                                |
| 100             | \$298 | 1400                  | .21286            |                                |
| 101             | \$300 | 1410                  | .21277            | $(300-298)/(1410-1400) = .2$   |
| 102             | \$302 | 1419                  | .21283            | $(302-300)/(1419-1410) = .222$ |
| ...             |       |                       |                   |                                |
| 200             | \$498 | 2100                  | .2371             |                                |
| 201             | \$500 | 2107                  | .2373             | $(502-498)/(2107-2100) = .286$ |
| 202             | \$502 | 2113                  | .2375             | $(502-500)/(2113-2107) = .333$ |

The overall cost-effectiveness ratio and marginal cost-effectiveness ratio for Health Intervention B is shown graphically in Figure 4-2. The fact that the marginal cost-effectiveness ratio curve intersects the overall cost-effectiveness ratio curve at the bottom of the latter curve is not coincidence. Whenever the overall cost-effectiveness curve is "U-shaped," as is the case here, the marginal cost-effectiveness ratio curve will intersect it at the bottom of the "U".<sup>8</sup>

What is the cost-effective level of Health Intervention B if its overall and marginal cost-effectiveness ratios look as they do in Figure 4-2? From Table 4-2, we would choose a cost level of \$500, serving 201 patients. For any incremental expansion of Health Intervention B at a level of size less than \$500, the additional expenditure brings in a greater level of health benefits than required by the threshold cost-effectiveness ratio. For any incremental expansion of Health Intervention B beyond size \$500, the additional expenditure results in a lower level of additional health benefits than required by the threshold cost-effectiveness ratio. Additionally, at size \$500 the overall cost-effectiveness ratio is .2373, well below the threshold cost-effectiveness ratio, indicating that a level of Health Intervention B of \$500 is indeed preferred to the status quo from a cost-effectiveness perspective. Thus in this case we would choose the optimal level of Health Intervention B just as we had in the prior example, by finding that point at which the marginal cost-effectiveness ratio equals the threshold cost-effectiveness ratio.

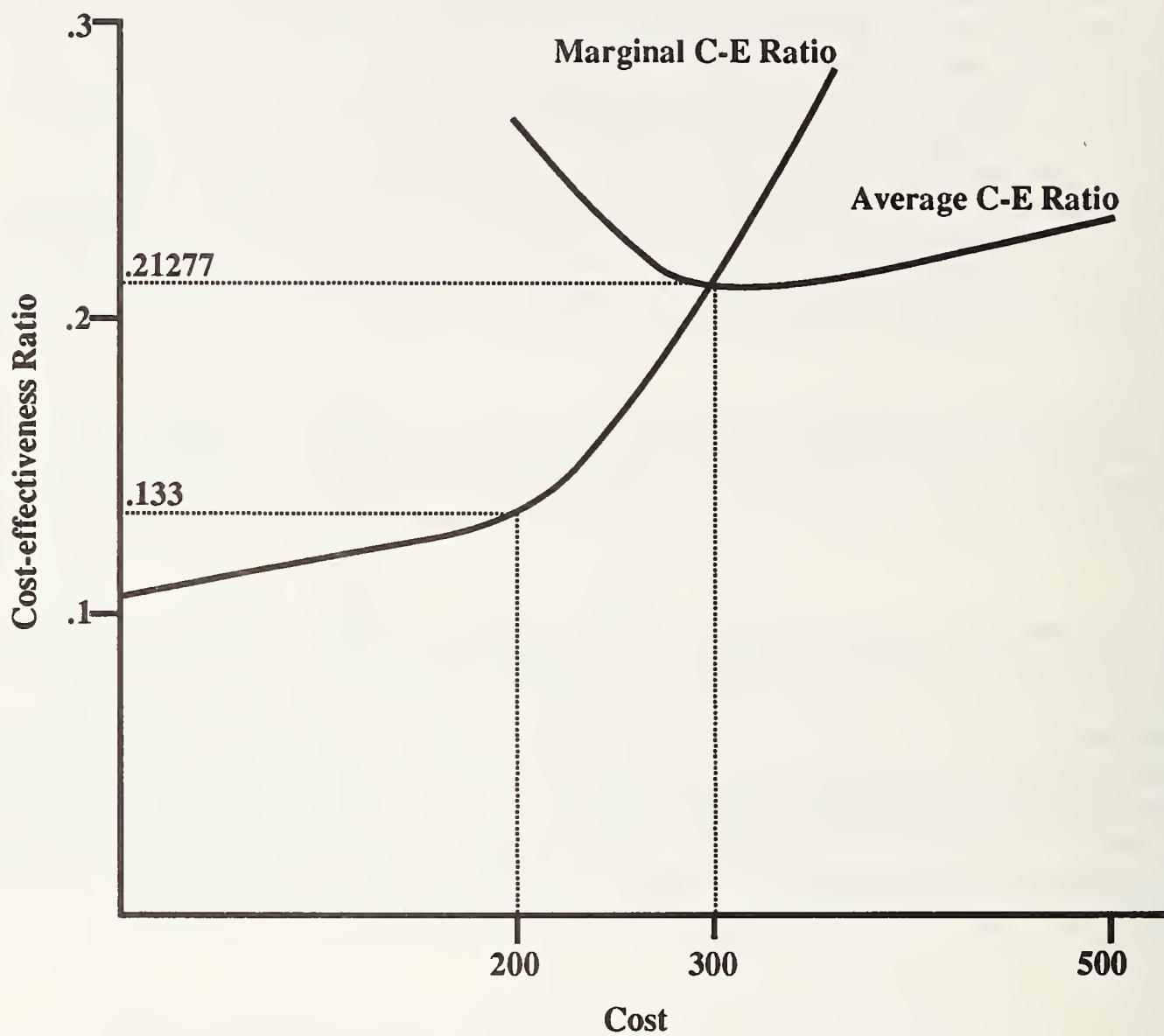
But, suppose the threshold cost-effectiveness ratio is .133. At a size of \$200, serving 51 patients, the marginal cost-effectiveness ratio is .133. But now a complication arises. The overall cost-effectiveness ratio at size \$200 is .248. This is above the threshold cost-effectiveness ratio. As such, we would conclude that Health Intervention B at a size of \$200 is not preferred to the status quo.

What has gone wrong in this last situation? Actually, the result is not as difficult to interpret as it may seem. The first patient required a very large expenditure. Thus, for the first dollars expended there is a very low return in terms of health. Subsequent expenditures result in much greater incremental health returns per dollar. Still, it may be that these subsequent large returns of health for incremental expenditure do not make up for the very low initial return of health for the first \$100. As such, it is the case that if the threshold ratio is .133, the expenditures from \$100 through \$200 are well worth the expense in terms of health returns per dollar expended, but they do not compensate for the very low return of the first \$100 in expenditure. Thus, if the threshold cost-effectiveness ratio is .133 it is better not to invest in Health Intervention B at all than to do so at size \$200.

---

<sup>8</sup>One way to see why the curves intersect as they do is to note that if the marginal cost-effectiveness curve is below the overall cost-effectiveness curve then additional expansion of the health intervention will involve less additional expenditure per unit of health benefit than the average up to that point. As such, additional expenditures will continue to lower the overall cost-effectiveness ratio. The logic is the same as if a baseball player goes into a game batting .333 and goes one for five, for a batting average of .200 for the day, his batting average after the game will be below .333. Conversely, if the marginal cost-effectiveness curve is above the overall cost-effectiveness curve then additional expansion of the health intervention will involve more additional expenditure per unit of health benefit than the average up to that point. As such, additional expenditures will increase the overall cost-effectiveness ratio. Thus, if the marginal cost-effectiveness ratio curve is below the overall cost-effectiveness curve when the latter is falling and is above the overall cost-effectiveness curve when the latter is rising, it must be that the switch from the marginal curve being below the overall curve to being above the overall curve occurs at the bottom of the overall curve.

Figure 4-2. The Cost-Effectiveness and Marginal Cost-Effectiveness of Intervention B



This last example illustrates the fact that while the condition  $MB/MC = R_{\text{Threshold}}$  is a necessary condition for a cost-effective level of a health intervention, it is not a sufficient condition. As in the cited example, it may be that  $MC/MB = R_{\text{Threshold}}$  but that  $TC/TB > R_{\text{Threshold}}$ . In that case, the intervention is not cost effective at that scale.

Or, it may even be that there are several levels of scale or size at which  $MC = MB = R_{\text{Threshold}}$ . This could happen if the total benefit curve is not concave but has sections that are concave and sections that are convex. The marginal benefit curve can decline in some regions, then begin to rise, then begin to decline again, and so forth. This can lead to several scale levels at which  $MC = MB = R_{\text{Threshold}}$ .

#### 4. A GENERAL ALGORITHM

Fortunately, an algorithm can be devised that covers all of the situations covered so far in this section:

- Find that level or levels of the health intervention at which the marginal cost-effectiveness ratio equals the threshold level.
- Check to make sure the overall cost-effectiveness ratio at this level (or these levels) is below the threshold level.
- If there is no such level meeting these above two stipulations, then it is cost effective not to invest in the health intervention at any scale.
- If there is more than one level meeting these two stipulations, choose among them in exactly the same fashion as when choosing among mutually exclusive health interventions.

Again, while this set of rules may seem a bit complex, in fact it is straightforward to apply. Moreover, it does not go against the basic definition of CUA as relying on (overall) cost-effectiveness ratios to determine if an intervention is preferable to the status quo. That rule remains firmly in place.

#### 5. IF "NOTHING" IS BEST, THE NEXT BEST MAY NOT BE "NEXT TO NOTHING"

There may be instances in which there are political or other forces dictating that a given health intervention be launched at some scale. That is, the option of remaining with the status quo may not be practical. But what if the cost-effective level of operation is at size zero? One might think that it is then best from a cost-effectiveness perspective to have the health intervention be as small as possible. After all, if the best level from a cost-effectiveness perspective is zero, does it not follow that the next best level is as close to zero as possible? The answer is that this is not always so.

Consider again the situation drawn in Figure 4-2. Suppose the threshold cost-effectiveness ratio is .133. As we indicated, in this case it is most cost effective for the intervention not to be undertaken at all. Nonetheless, if it must be undertaken, the next best situation is to operate up to a level of 51 patients. This is because, once the first patient is served and the initial cost of \$100 is incurred, it is cost effective to serve each of the subsequent patients, at least until 51 patients are served. For each of these subsequent patients, the incremental health benefits are worth the additional expense. This ceases to be true only after patient 51, when the marginal cost-effectiveness ratio rises above .133.

This type of example provides an important lesson in many real situations in which, for reasons which may or may not make economic sense, some large facility or expensive new piece of equipment is put into place. Even if the most cost-effective situation would have been not to have the facility or new piece of equipment in the first place, the next best thing from a cost-effectiveness perspective may well be to try to exploit it extensively.

## 6. MEASURING SCALE

For simplicity, it was assumed in the previous discussion that the scale of a health intervention can be captured by its cost. This is a natural measure of scale if there is some single physical dimension that characterizes the size of an intervention, e.g., number of rooms in a hospital, amount of medication, number of visits to a doctor, etc. In such cases, scale and cost move together in an unambiguous fashion.

But what if there is no one single dimension of scale but, rather, several that operate simultaneously? For instance, a hospital's scale may be characterized not only by the number of rooms but by the equipment available, the types of operating rooms, and so forth. To handle these cases, let  $X_1, X_2, \dots, X_K$  be the different dimensions of scale that are relevant. Total costs will be a function of the value of each of these dimensions, as will total health benefits. This can be denoted by writing TC and TB as  $TC(X_1, X_2, \dots, X_K)$  and  $TB(X_1, X_2, \dots, X_K)$ .

A cost-effectiveness ratio can be defined for any given value for  $X_1, X_2, \dots, X_K$  as  $TC(X_1, X_2, \dots, X_K)/TB(X_1, X_2, \dots, X_K)$ . Similarly, a marginal cost-effectiveness ratio for any given value for  $X_1, X_2, \dots, X_K$  and for *each* scale dimension  $i$  can be defined as:  

$$MCER_i = [\partial TC / \partial X_i] / [\partial TB / \partial X_i].$$

The numerator of this ratio,  $\partial TC / \partial X_i$  indicates the amount by which total cost, TC, changes when there is a one unit increase in the value of the  $i$ th dimension of scale. Similarly, the denominator,  $\partial TB / \partial X_i$ , indicates the amount that total benefits, TB, changes when there is a one unit increase in the value of the  $i$ th dimension of scale. The ratio of these two terms thus represents the "marginal cost-effectiveness ratio," i.e., the change in costs divided by the change in benefits that result from a one unit change in scale dimension  $i$ .

Now the logic of the prior subsections can be used to choose the cost-effective scale size for each dimension of scale. The sufficient condition for a cost effective scale for scale dimension  $i$  becomes  $MCER_i = R_{\text{Threshold}}$ . As above, while a sufficient condition, this is not a necessary condition. Analogous to what must be done above, even if  $MCER_i = R_{\text{Threshold}}$  for each  $i$ , one must check that at the associated scale TC/TB is also less than  $R_{\text{Threshold}}$ .

## 7. MIXING THE "WHICH" AND THE "HOW MUCH" QUESTIONS

In this chapter, we have analyzed the "how much" question, looking at the cost-effective size of a health intervention, while the previous chapter looked at choosing among mutually exclusive health interventions. What if the situation were a combination of the two?

Suppose a choice must be made among several mutually exclusive health interventions, each of which can be pursued at different scales or sizes of operation.

This is straightforward. For each intervention, determine the optimal size from a cost-effectiveness perspective, treating that intervention in isolation. Then choose among the health interventions as a set of mutually exclusive interventions, each pegged at the level which is most cost effective for the intervention when considered by itself.



## CHAPTER 5

### OTHER CONSIDERATIONS IN APPLYING THE MECHANICS OF CUA

#### 1. INTRODUCTION

Up to now we have assumed that health interventions cost Government resources and provide health benefits—in other words, that both numerator and denominator of the cost-effectiveness ratio are positive. This chapter discusses three cases in which this assumption does not hold and provides a modified algorithm to deal with these situations. The chapter concludes with a discussion of the nature of the status quo in cost-utility analysis (CUA).

#### 2. WHAT IF THE NUMERATOR AND DENOMINATOR IN THE COST-EFFECTIVENESS RATIO ARE NOT BOTH POSITIVE?

##### 2.1. Programs That Save Money and Improve Health

A health intervention may provide health benefits and save resources at the same time. Anderson, et al. (1986) argue that:

...actions that generate both positive outcomes and positive economic returns are rare in business or economics. They are common in health, however, and since the technical interpretive problem sometimes causes confusion, a general solution is desirable.<sup>9</sup>

In such cases, the same basic logic applies as used throughout. The cost-effectiveness ratio is now negative, however, since the numerator—costs of the health intervention—is a negative number, reflecting cost savings, while the denominator—health benefits—is positive. As long as the threshold cost-effectiveness ratio is positive, as it invariably will be, the cost-effectiveness ratio of the health intervention under consideration will be below the threshold level; compared with the status quo, health interventions that provide health benefits and save money at the same time are clearly cost-effective. Thus the basic algorithm for determining if an intervention is preferred to the status quo remains valid.

Choosing among mutually exclusive health interventions that save money involves the same algorithm used when choosing among mutually exclusive health interventions. The health interventions should be ordered from the smallest to largest in cost. Now the smallest in cost is the intervention that saves the most money. As one moves down the list from this health intervention that saves the most money to that which saves the next most money, there is a net cost equal to the difference in savings between the two health interventions.

The question then becomes whether there are additional health benefits to justify this additional cost. This question can be answered using the incremental cost-effectiveness ratio, outlined in Chapter 3. Indeed, all of the analysis of Chapter 3 remains valid for choosing among mutually exclusive interventions, each of which or some of which provide health benefits while at the same time saving money.

---

<sup>9</sup>John P. Anderson, J.W. Bush, Milton Chen, and Danielle Dolenc, Policy Space Areas and Properties of Benefit-Cost/Utility Analysis, *Journal of the American Medical Association*, 255(4):795, 1986.

## 2.2. Programs That Save Money But Reduce Health

Once one leaves the benevolent world of health interventions that simultaneously save money and improve health, there remain other situations in which either the numerator or denominator of a cost-effectiveness ratio is not positive. One of them is when a health intervention that costs money but improves health is discontinued. This action of discontinuing the program can itself be thought of as a health program, one that saves money but lowers health as a result.

Consider a health intervention, Health Intervention A, that saves \$100 and lowers health by five units. The cost-effectiveness ratio is  $-\$100/-5$ , or 20. Both the numerator and denominator are negative, making the ratio positive. If  $R_{\text{Threshold}}$  is 10, then using our prior algorithm, Health Intervention A would not be judged to be cost-effective; the status quo would appear preferable. But now imagine having adopted Health Intervention A, and consider "returning" to the old status quo. Health Intervention Old Status Quo (OSQ) would lead to a cost of \$100 and a health improvement of five units, for a cost-effectiveness ratio of, again, 20. Now, Health Intervention A is preferred to the Health Intervention OSQ.

Which of these contradictory conclusions is correct? The latter one is correct. Health Intervention A is preferred to Health Intervention OSQ. This is because when both the numerator and denominator of a cost-effectiveness ratio are negative, it is no longer the case that the lower the ratio the more desirable the intervention. Just the opposite is now true. If a health intervention lowers health by a given amount and saves money, clearly the more money saved (holding the deterioration in health fixed) the better. The higher the cost-effectiveness ratio the better.

## 2.3. Programs That Cost Money and Worsen Health

The third and final situation we consider in which the numerator and denominator are not both positive is when an intervention costs money but worsens health. There are instances of iatrogenic illnesses in which, despite the best of intentions, an intervention has led to worse levels of health while costing resources.<sup>10</sup> These types of health interventions have a positive numerator and a negative denominator in their cost-effectiveness ratios. Again, we cannot use our traditional guideline for determining if a given intervention of this type is cost effective compared to the status quo. Clearly an intervention that costs money and lowers health is not cost effective. But the cost-effectiveness ratio is negative and therefore will be below the threshold level indicating, incorrectly, that the intervention is cost effective.

## 2.4. A General Algorithm

Fortunately, all of these situations can be accounted for by a relatively minor modification to our basic algorithm.

- If the cost-effectiveness ratio of a health intervention is lower than the threshold level, then the intervention is cost-effective relative to the status quo *unless* the cost-effectiveness ratio is based on a negative denominator—i.e., the health intervention lowers health benefits.

---

<sup>10</sup>See, e.g., Robert F. Morgan, Ed., *The Iatrogenics Handbook: A Critical Look at Research and Practice in the Helping Professions*, Toronto, Ontario: IPI Publishing, 1983.

- If the cost-effectiveness ratio is based on a negative denominator *and* is above the threshold level, then the health intervention is cost effective relative to the status quo.

## 2.5. A Graphic Analysis

One way to get a better sense of the situations considered in this section is through a graphic analysis such as in Figure 5-1, adapted (with transposed axes) from Anderson, et al., (1986). The horizontal axis in Figure 5-1 represents the net cost of the program. Positive numbers on this axis mean the program costs money, negative numbers on this axis means the program saves money. The vertical axis represents the net change in aggregate health outcomes. (As throughout this chapter, let us not worry for now exactly how such health outcomes are measured.) Positive numbers on this axis mean the programs improve health while negative numbers mean the programs lead to worse health outcomes. Costs and health outcomes are measured relative to the status quo, which can be thought of as being located at the origin of the graph. Situations where a health intervention saves money and improves health are located in the fourth quadrant, on the upper left. Health interventions which save money but lead to lower health are located in the third quadrant, on the lower left. Health interventions which cost money and worsen health are located in the second quadrant, in the lower right. Finally, health interventions which cost money but improve health, as supposed in the analysis of the previous section, are located in the first quadrant, on the upper right.

It can be noted that in general the evaluation of health interventions relative to the status quo can be thought of as involving a tradeoff between health and costs. This occurs for health interventions located in either the first or third quadrant. In these cases, either improved health comes at a price—in which case the numerator and denominator of the CUA ratio are both positive and the intervention lies in Quadrant 1—or savings in costs lead to lower health outcomes—in which case the numerator and denominator of the CUA ratio are both negative and the intervention lies in Quadrant 3.

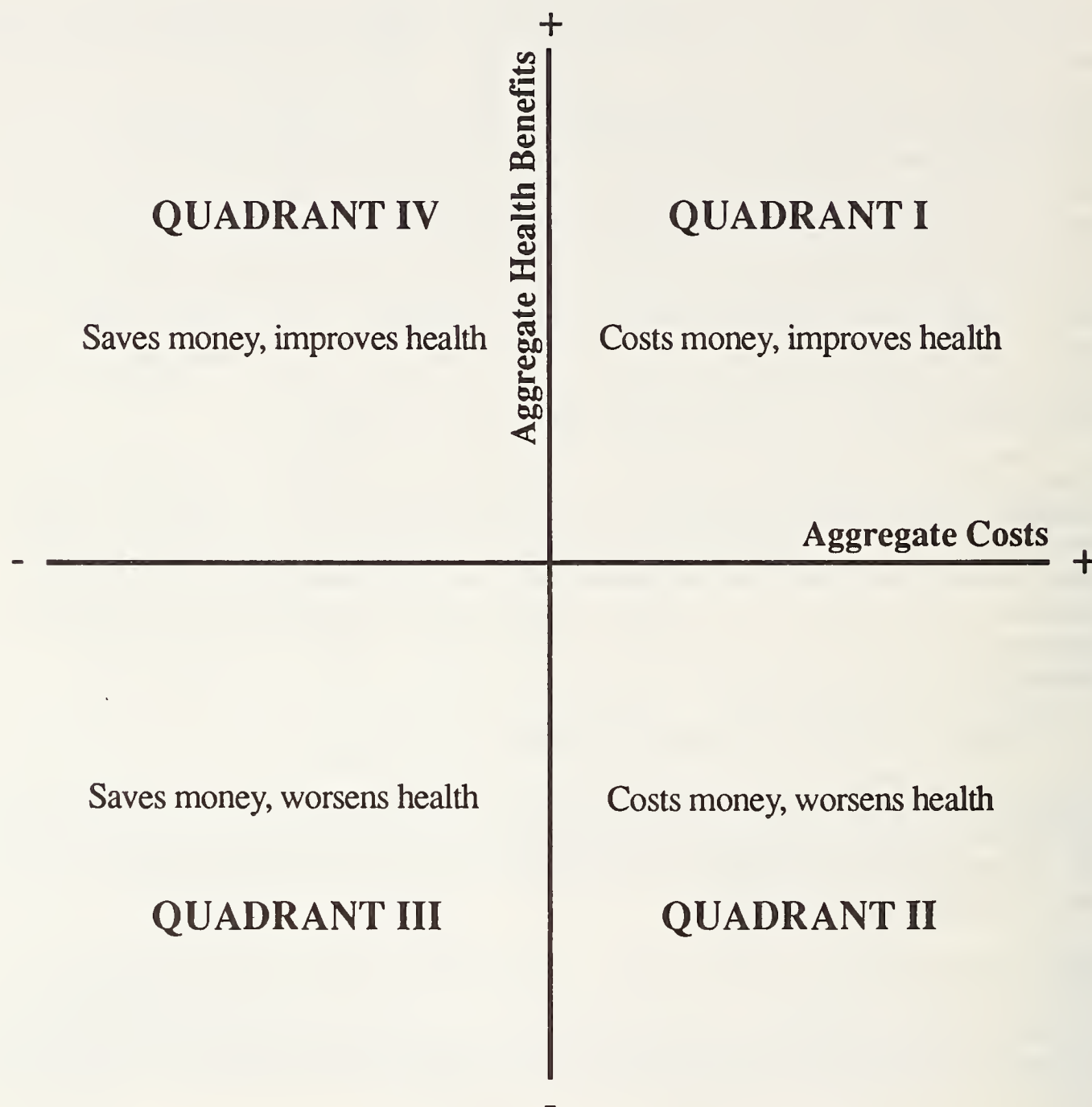
## 3. THERE IS NOTHING SPECIAL ABOUT THE STATUS QUO.

CUA has been defined as a way to determine the cost-effectiveness of a health intervention compared to the status quo. It may appear from this definition that CUA places undue emphasis on the status quo. It may also seem that the analysis assumes that the status quo represents a special or perhaps a particularly desirable state of affairs.

Neither of these inferences would be correct. CUA attaches no special status to the status quo. A health intervention is compared with the status quo in CUA only because if the health intervention is not adopted then the status quo is what remains. That is, evaluation of health interventions implicitly involves a comparison between that health intervention and its alternative, which is not adopting the intervention and therefore remaining with the current situation.

The status quo should not be confused with "doing nothing." In general, the status quo involves treating a patient in a particular way. CUA is best thought of as a method of evaluating potential health interventions which would be substitutes for regular treatment methods. Moreover, it does not matter to the outcome of a CUA what alternative is taken to be the status quo. Consider the task of

**Figure 5-1. Possible Configurations for the Relationship of Aggregate Costs to Aggregate Benefits**



choosing among mutually compatible interventions. Suppose one relabelled the status quo as Health Intervention 0 and took some other health intervention and called it the status quo. All of the health interventions would now have their costs and incremental health outcomes measured relative to the new status quo. Following the same algorithms previously put forth, all of the results of the analyses remain identical.

This can be illustrated using our example from Chapter 3 involving choice among mutually exclusive health interventions. The interventions considered there are summarized by the following table showing costs and health outcomes relative to the status quo.

**Table 8. Mutually Exclusive Health Interventions**

| Health Intervention | Cost | Change in Health Status |
|---------------------|------|-------------------------|
| 0 (status quo)      | \$ 0 | 0                       |
| 1                   | 30   | 30                      |
| 2                   | 30   | 20                      |
| 3                   | 55   | 70                      |
| 4                   | 20   | 40                      |
| 5                   | 30   | 50                      |
| 6                   | 50   | 75                      |

Let the status quo now be redefined as, say, Health Intervention 2. Remeasuring costs and the change in health status relative to Health Intervention 2 gives the following table, ordered by cost, as specified by step 1 of the first algorithm in Chapter 3 for choosing among mutually exclusive health interventions (Table 9).

**Table 9. Mutually Exclusive Health Interventions  
Measured Relative to a New Status Quo**

| Health Intervention | Cost | Change in Health Status |
|---------------------|------|-------------------------|
| 0 (old status quo)  | -30  | -20                     |
| 4                   | -10  | 20                      |
| 2 (old status quo)  | 0    | 0                       |
| 1                   | 0    | 10                      |
| 5                   | 0    | 30                      |
| 6                   | 20   | 55                      |
| 3                   | 25   | 50                      |

Following step 2 of the algorithm to examine for cases of strong dominance, Health Intervention 2, the new status quo, and Health Intervention 1 are dominated by Health Intervention 4, which costs less (saves money relative to Health Intervention 1 or 2) and provides fewer health benefits. Health Intervention 3 is dominated by Health Intervention 6 on the same grounds. Removing these health interventions yields the following table, with the indicated incremental cost-effectiveness ratios.

**Table 10. Eliminating Strongly Dominated Health Interventions**

| Health Intervention | Cost | Change in Health Status | Incremental CE Ratio |
|---------------------|------|-------------------------|----------------------|
| 0 (old status quo)  | -30  | -20                     |                      |
| 4                   | -10  | 20                      | .5                   |
| 5                   | 0    | 30                      | 1.0                  |
| 6                   | 20   | 55                      | .876                 |

Applying step 4 of the algorithm to identify weakly dominated health interventions indicates that Health Intervention 5 is weakly dominated by Health Intervention 6. Removing Health Intervention 5

from the list, though, leaves us with the exact same table that we faced following step 4 in Chapter 3. Identical to that analysis, if the threshold cost-effectiveness ratio is less than .5 then Health Intervention 0, the old status quo, should be chosen. If the threshold cost-effectiveness ratio is between .5 and .876, then Health Intervention 4 should be pursued. And, if the threshold cost-effectiveness ratio is greater than .876 then Health Intervention 6 should be pursued. These are exactly the same conclusions we reached in Chapter 3.



## PART II

### NORMATIVE UNDERPINNINGS TO CUA

Ethical norms and values acceptable to a society must ultimately be at the core of any theory of resource allocation in health care (especially as pursued by the public sector). Without such grounding, a theory of resource allocation lacks any fundamental justification or rationale. However, most works in cost-effectiveness analysis in health care steer clear of such concerns, in part no doubt because of the difficulty of the issues involved and the debate and controversy they may engender. Sometimes this comes at little cost. Sometimes common sense alone is sufficient to answer measurement and interpretation questions without having to deal explicitly with foundational questions. But presentations of CUA in health care that ignore ethical and normative underpinnings do so ultimately at their own peril. In difficult areas of interpretation or measurement, one is left with no basis upon which to make judgments unless a coherent normative framework is advanced.

Allan Gibbard argues the case when he writes:

In many realms of life, no doubt, satisfactory ethical judgments can be made without careful analysis of the fundamental bases of these judgments. That is unlikely to be the case, however, with questions of health policy. When applied to those questions, whatever "moral good sense" we may develop in the ordinary course of life is likely to be inadequate. The effects of health policy are immensely complex, and so we cannot simply take in the nature and the effects of a set of policies at a glance and focus our moral good sense on them....To make well-founded moral judgments on the basis of sophisticated analysis of the effects of policy, we shall need sophisticated ethical theory. We shall need to work out explicitly the kinds of ethical considerations we think should govern and the weights these considerations should have.<sup>11</sup>

Any specific model or set of procedures for CUA in health care implies ethical norms. It can be argued that the only issue in presenting a set of rules and procedures for CUA is whether the underlying ethical norms are implicit, and possibly murky and confused, or explicit and open to clear debate and analysis. The bias presented here is, for the latter, in concurrence with Emery and Schneiderman (1989) in their discussion of CEA:

Far from being a pristine mathematical algorithm, CEA employs complex and often overlooked value components along with empirical data...The more widely CEA is utilized, the greater is the need to recognize and clarify the ethical concerns it raises.<sup>12</sup>

In considering the normative basis for CUA, two basic questions arise repeatedly:

---

<sup>11</sup>Allan Gibbard, *The Prospective Pareto Principle and Equity of Access to Health Care*, Appendix G of *Securing Access to Health Care: The Ethical Implications of Differences in the Availability of Health Services*, Volume Two: Appendices, President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, Washington, DC: U.S. Government Printing Office, 1983, p. 155.

<sup>12</sup>Danielle Colence Emery and Lawrence J. Schneiderman, *Cost-effectiveness Analysis in Health Care*, *Hastings Center Report*, July/August 1989, p. 8.

**Interpretive Question 1:** *What does it mean to say that when an individual moves from Health State C to Health State A his or her health improves "twice as much" as moving from Health State C to Health State B?*

**Interpretive Question 2:** *Why is an improvement by one individual by 15 health units "better" than an improvement by two individuals of five health units each?*

Both of these types of judgements are required by CUA as defined here. But on what basis might such judgements conceivably be made? In the context of the first interpretive question, what does it really mean to say that one change in a person's health state represents an improvement that is twice (or three times or half) as much as some other change in his health state? In the context of the second interpretive question, how can an intervention that helps one individual by a lot be justified in place of another intervention that helps two individuals by a lesser amount?

As will be seen, the extent to which CUA can serve as a compelling guide to resource allocation in health care is in no small part a function of how convincingly and coherently these questions can be answered. As such, in presenting a normative framework for CUA, the following chapters can be viewed as essays designed to address these questions.

## CHAPTER 6

### CUA-I: THE INDIVIDUAL PERSPECTIVE AND THE BASICS OF VNM-UTILITY THEORY

#### 1. INTRODUCTION

CUA-I considers cost-utility analysis (CUA) from the perspective of an individual decisionmaker. CUA-I is designed to help an individual decide, for example, whether to have by-pass surgery or not, how frequently to have mammograms, or whether to take medicine X or Y for a particular illness. (Alternatively, and without too much substantive distortion to CUA-I, it may be viewed as a guide to physicians and other health care professionals in terms of what to recommend to their patients.)

On what basis should an individual make decisions about which health care interventions to adopt? Often, this is a decision that must be made under substantial uncertainty since the health outcomes of a given intervention generally are not completely known in advance. Additionally, most such decisions involve tradeoffs between health effects (or the probability of health effects) and monetary as well as other types of costs.

One of the most powerful normative frameworks for determining how an individual should make such decisions is *utility theory*. This chapter introduces utility theory and indicates the link between CUA-I and utility theory. It does so assuming that the only costs associated with a health intervention are health care resources and that the only impacts are on the health of the individual involved. Subsequent chapters consider modifications necessary to this basic framework when the situation is more complicated.

#### 2. BASICS OF UTILITY THEORY: THE EXPECTED UTILITY THEOREM

Utility theory in its modern form originated with VonNeumann and Morgenstern (1947). To distinguish this form of utility theory from other, more casual uses of the term "utility," it will sometimes be referred to as VNM-utility. VNM-utility theory has had a major influence in many areas of the social sciences and in policy analysis. It has served as the foundation for incorporating uncertainty into neoclassical economic theory and it is the basis for much of decision analysis, multiattribute utility theory, and game theory.<sup>13</sup> It is also of central relevance for CUA in health care.

Consider an individual facing several mutually exclusive health interventions, each of which costs the same amount. The outcome of some of the interventions may be uncertain. Which health intervention should the individual choose? From the perspective of utility theory, the following procedure should be used.

---

<sup>13</sup>Classic introductions to decision analysis, decision analysis as applied to medical decision making, and multiattribute utility theory are, respectively, Howard Raiffa, *Decision Analysis: Introductory Lectures in Choice Under Uncertainty*, Addison-Wesley, Reading, Mass., 1968; Milton Weinstein and Harvey Fineberg, *Clinical Decision Analysis*, W.B. Saunders, Philadelphia, 1980; and Ralph Keeney and Howard Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley and Sons, New York, 1976.

First, align all the health states that might result from any intervention in order from best to worst. Let Health State A be the best health state, Health State B the next best, and so on, up to, say, Health State Z. Assign Health State A a utility value of 1 and Health State Z a utility value of 0.<sup>14</sup>

Now consider some health state between these two, Health State D, which is better than Health State Z and worse than Health State A. Ask the individual this question: "You face the following choice. You can either be in Health State D for sure, or you can play a game of chance in the form of a lottery. The lottery assigns you either to Health State A or Health State Z. It assigns you to Health State A with a probability of  $p$  and to Health State Z with a probability of  $1-p$ . What is the value of  $p$  over which you would choose the lottery and under which you would choose Health State D? Put another way, what is the value of  $p$  for which you are indifferent between playing the lottery and being in Health State D with certainty?" If  $p$  is very near 1 then the lottery basically amounts to being in Health State A, so the lottery will be preferred. If  $p$  is very near 0, then the lottery basically amounts to being in Health State Z, so the lottery will not be chosen. There must be some "cutoff" point, however, at which, if  $p$  is greater than the cutoff, the lottery will be chosen while if  $p$  is less than the cutoff, Health State D will be chosen instead.

Making a judgement about the cutoff value for  $p$  may be very difficult. Moreover, the decision about the cutoff value for  $p$  may vary between individuals. But upon reflection, most individuals should be able to make a judgement about  $p$  for most sets of health states that might be considered. This value of  $p$  for Health State D can be termed the VNM-utility of Health State D. In terms of functional notation, we then have:

$$(6-1) \quad U(D) = p_D,$$

where  $U(D)$  means the "VNM-utility of Health State D." A similar exercise can be conducted for all of the other health states so that each health state has a VNM-utility.

Why go through such an odd exercise? The answer is provided by the so-called *expected utility theorem*. This theorem states: *in choosing among different options involving different outcomes (in this case, different health interventions involving different health outcomes), an individual should take the action that maximizes the expected utility of the outcome.*

To compute the expected utility of a health intervention, each possible outcome from the intervention should be considered. The VNM-utility from each outcome is then multiplied by the probability of that outcome and the resulting figures for each outcome added together. This is illustrated below.

The expected utility theorem is based on a set of axioms of individual choice. While these axioms of choice are not examined in detail here, they are widely viewed as normatively compelling and embodying "common sense," with any deviation from them generally seen as a "mistake" which the individual, upon reflection, would wish to have rectified. Often, economists (and others) describe decisionmaking that is consistent with these axioms and therefore consistent with the expected utility theorem as "rational."

---

<sup>14</sup>It turns out these numerical assignments of 1 and 0 are not necessary. Health State A could be assigned a value of 37 and Health State Z a value of -22 and the logic of the approach would remain intact. However, the 1/0 numerical assignment makes for easier exposition.

The expected utility theorem indicates how someone should make decisions in choosing among alternatives when the outcomes of a given choice are not known with certainty. The expected utility theorem is a normative statement and not necessarily a positive or behavioral statement. That is, it is not necessarily claimed on the basis of experimentation that individuals for some reason or other happen always to make choices precisely in this fashion. The theorem is prescriptive, not necessarily descriptive.

### 3. AN ILLUSTRATION

To illustrate the application of utility theory to health care, consider an example. Suppose an individual develops a rash. Two treatments are available, both of which cost the same. Treatment 1 cures the rash in 2 weeks for all patients. There are no side effects. Treatment 2 has a less predictable outcome. In 80 percent of the cases, the rash goes away within 1 day (with no side effects). In 15 percent of the cases, the rash goes away in 2 weeks (with no side effects). In the remaining 5 percent of the cases, there is an adverse reaction (high fever), and the individual must spend about a week in bed, although the rash goes away in 2 weeks. Which treatment is appropriate? The choice between Treatment 1 and Treatment 2 seems far from straightforward even though the example has been constructed so there are no financial tradeoffs to consider. Utility theory provides a method for the individual to make the choice "rationally." It does not prejudge the choice or indicate that the decision so reached should be the same for everyone. But, it does provide an operation method for a given individual to make such a choice.

In this example there are three outcomes: Outcome A (the best): rash goes away in 1 day, no side effects; Outcome B (middle): rash goes away in 2 weeks, no side effects; and Outcome C (the worst): rash goes away in 2 weeks but with side effects. Assign Outcome A a utility value of 1 and Outcome C a utility value of 0. Now, have the individual consider the choice between having Outcome B with certainty and having a "lottery ticket" that yields Outcome A with probability  $p$  and Outcome C with probability  $1-p$ .

At what value of  $p$  is the individual indifferent between the two options? Let this value be denoted as  $U(B)$ , indicating that this is the VNM-utility assigned to Outcome B. As suggested earlier, this choice about  $p$  may not be easy for an individual to reach but, at the same time, it should not be impossible given sufficient reflection.

Having determined  $U(B)$ , the individual's basic decision problem can now be solved. The utility of Treatment 1 is  $U(B)$ . This outcome is obtained with certainty if Treatment 1 is adopted, so the expected utility of Treatment 1 is  $(1 \times U(B))$  or  $U(B)$ . Treatment 2 leads to Outcome A (with a utility of 1) with probability .8, Outcome B (with utility of  $U(B)$ ) with probability .15, and Outcome C (with utility of 0) with probability .15. So, the expected utility of Treatment 2 is  $(1 \times .8) + (.15 \times U(B)) + (.05 \times 0)$ , or  $.8 + .15U(B)$ . Based on the expected utility theorem, Treatment 1 is preferred if its expected utility,  $U(B)$ , is greater than the expected utility of Treatment 2,  $.8 + .15U(B)$ . Or, doing some basic algebra, Treatment 1 is preferred to Treatment 2 if  $U(B)$  is greater than .941 (rounding off to three digits).

It may seem that this result is magical. Why is .941 the dividing point? Nonetheless, as long as one accepts the basic axioms underlying the expected utility theorem, the normative validity of which have been almost universally recognized, this result follows as a matter of logic. The fact that the expected utility theorem allows decisions such as those in this example to be "solved" is not magic but testimony to the power of the approach and its potential for use in health policymaking.

#### 4. LINKING VNM-UTILITY THEORY TO CUA

The previous discussion avoided tradeoffs between money and health to avoid introducing too many complications at once. However, by avoiding these tradeoffs, which are central to CUA, the link between the prior discussion and CUA remains a bit obscure. Now let's turn to the issue of tradeoffs between money and health.

Suppose an individual faces various mutually compatible health interventions. Each health intervention,  $HI_i$ , costs some amount of money,  $C_i$ , and leads to some change in health,  $\Delta h_i$ . Let  $\Delta h_i$  be a description of a health change itself, not necessarily a numerical measure. An example of  $\Delta h_i$  would be "prevents me from getting a cold" or "gets rid of my back pain."<sup>15</sup>

A given health intervention may not lead deterministically to a known health outcome. In that case, a set of changes in health states can be identified,  $\Delta h_{i,1}, \Delta h_{i,2}, \dots, \Delta h_{i,n}$ , which might result from a given health intervention  $i$ . Let the probability that health intervention  $i$  leads to health outcome  $\Delta h_{i,n}$  be denoted by  $p_{i,n}$ . One may view  $p_{i,n}$  as an "objective" probability, based, for example, on controlled clinical trials. Or one may view  $p_{i,n}$  from a more Bayesian<sup>16</sup> perspective as a subjective probability, based on the individual's (or his doctor's) evaluation of the likelihood of the given health change resulting from that intervention.

Suppose that an individual has a budget constraint defined in terms of the amount of financial resources he can spend on a set of mutually compatible health interventions. This will not alter the fact that he will wish to make his choices consistent with the expected utility theorem. Now, however, he will wish to maximize his expected utility subject to this budget constraint.

It is here that the link between CUA-I and utility theory is forged. As seen in Chapter 3, the basic definition of CUA is mathematically equivalent to maximizing aggregate health outcomes subject to a budget constraint. In the world of CUA-I there are no aggregate health benefits, only individual benefits. So, the basic definition of CUA is mathematically equivalent to maximizing individual health benefits subject to a budget constraint. Substituting "the expected utility of individual health benefits" for "individual health benefits" in this definition of CUA yields the individual's goal as specified in the previous paragraph. That is, normative principles of rational choice as put forth by utility theory will lead an individual to adopt CUA in the form presented here, where health is measured in terms of VNM-utility. The rules of CUA are not arbitrary but can be derived from their normative foundations as specified by the principles of utility theory and decision theory!

Thus, to maximize the expected VNM-utility of health changes subject to a budget constraint, the individual will use  $\sum p_{i,n} U(\Delta h_{i,n})$  as a measure of the health outcome for each health intervention  $i$

---

<sup>15</sup>The description of the health state must also indicate the time at which it occurs, its duration, severity, and so forth. For simplicity of exposition, we avoid such a full-scale description here.

<sup>16</sup>"Bayesian" refers here to an approach to statistical decision theory that allows the probability of an event to be based on the considered judgment of an individual about the likelihood of the event occurring. This probability judgment is a function of the information that the individual making the judgment has at the time the judgment is made. Bayesian analysis also devotes attention to how an individual's initial judgment about the likelihood of an event, his "prior probability distribution," should be modified to form his "posterior probability distribution" when new information is received.

under consideration.<sup>17</sup> Using this as a health outcome measure, the CUA rules put forth in Chapters 2 through 5 for choosing a set of health care interventions will be applied. Just as specified in Chapter 2, a threshold for the ratio of costs to the expected utility of health changes from a given intervention,  $R_{\text{Threshold}}$ , can be defined. Health interventions with cost-effectiveness ratios below this threshold will be adopted by the individual—they will be cost effective— while those with cost-effectiveness ratios above this threshold will not be adopted.

## 5. ADDRESSING THE INTERPRETIVE QUESTIONS

The introduction to Part II presented two questions of interpretation, which any measurement procedure in CUA had to address. The second of the two questions concerned comparing health benefits across individuals. In CUA-I this issue never arises since the only concern is with a given individual. There are no aggregate social health benefits to consider. Thus, from the CUA-I perspective, there is no need to address the second interpretive question.

The first interpretive question asked how one could judge whether the move from Health State C to Health State A improved health by a certain amount compared to the move from Health State C to Health State B. In the context of VNM-utility theory, comparisons between the health improvement of moving from Health State C to Health State A relative to moving from Health State C to Health State B concern the following ratio:

$$(6-2) \quad [U(A)-U(C)]/[U(B)-U(C)].$$

If this ratio is 2, for instance, then moving from Health State C to Health State A represents twice the health improvement of moving from Health State C to Health State B. (Or, equivalently, Health State B is exactly half way between Health States C and A in terms of healthiness.)

The illustration from Section 3 provides an example of the role of this type of judgment. Recall that in the illustration there were two mutually exclusive health interventions, 1 and 2, and three health states, A, B, and C. The conclusion of the analysis was that, using the value of 1 for  $U(A)$ , and the value of 0 for  $U(C)$ , then Health Intervention 1 is preferred if the utility of Health State B is greater than .941.

Assume  $U(B)$  is equal to this threshold value of .941. Then, the ratio represented in Equation (6-2) is  $[1 - 0]/[.941 - 0] = 1.063$ . If  $U(B)$  is greater (less) than .941, then this ratio will be less (greater) than 1.063. So, the conclusion about which health state is preferred can be restated as follows: Health Intervention 2 is preferred only if the move from Health State C to Health State A improves health by an amount at least 1.063 as much as (or, equivalently, 6.3 percent greater than) moving from Health State C to Health State B.

---

<sup>17</sup>Here and throughout we use the notation  $U(\Delta h_{i,n})$  as short-hand for  $U(h_{i,n}) - U(h_{i,0})$ , where  $h_{i,n}$  is the health state of the individual after the health intervention and  $h_{i,0}$  is the health state of the individual in the status quo situation. If the potential for multiple possible health outcomes from health intervention  $i$  is not being emphasized, the notation  $U(\Delta h_i)$  will often be used, where  $U(\Delta h_i)$  is short-hand for  $U(h_i) - U(h_0)$ .

This notation should be contrasted with that in Chapter 5. There,  $\Delta H_i$  represented a numerical measure of the change in (aggregate) health from a health intervention. That was before the notion of VNM-utility was introduced. Now, we distinguish between the description of a change in health brought about by a health intervention,  $\Delta h_i$ , and its numerical measure,  $U(\Delta h_i)$ , where both  $\Delta h_i$  and  $U(\Delta h_i)$  refer to an individual health outcome.

The ability to express the judgment between Health Interventions 1 and 2 in this fashion reflects our conclusion that CUA requires precise numerical comparisons of the relative magnitude of different health improvements. But, what does it mean for the move from Health State C to Health State A to be twice as much as (or three times as much as, or 6.3 percent greater than) the move from Health State C to Health State B? Arithmetically, here it simply means that the values of  $U(A)$ ,  $U(B)$ , and  $U(C)$ , computed in the fashion prescribed, are such that the above ratio assumes a value that is 2 (or 3, or 1.063).

More substantively, though, we can say that this set of numerical values is made meaningful by VNM-utility theory itself. These numbers are the ones used in making the expected utility theory operational. The expected utility theory is, in turn, based on axioms of choice which are widely viewed as normatively compelling. Further, the choices about health interventions derived from these utility judgments can be determined using CUA as defined in Chapter 1 with the associated mechanics put forth in Chapters 2-5.

This is a strong and compelling answer to Interpretive Question 1 and one that provides a strong justification for CUA-I. It implies that the calibration procedure put forth by VNM-utility theory is the *only* measurement procedure that allows for "rational" decision making about health outcomes. Choice in any manner other than that prescribed would implicitly violate one or more of these basic axioms of rational individual choice, meaning that the conclusions would be inconsistent with the individual's own preferences and desires.

This chain of reasoning leads to the following two strong conclusions:

- CUA-I in the manner prescribed represents the proper procedure for an individual to choose among health interventions given a fixed budget to spend on health interventions. (The requirement that there be a budget constraint is relaxed subsequently.)
- VNM-utility assessments of health states represent the proper measurement of health status for CUA-I. Other procedures for assigning health states are valid only for CUA-I to the degree to which they are consistent with VNM utility.

## 6. SOME BASIC MULTIATTRIBUTE UTILITY NOTATION

So far it has been assumed in this chapter that health interventions consume health care resources, the magnitude of which can be represented in monetary costs, and that these interventions affect an individual's health. Under this assumption, the only relevant factors in considering health interventions are health and money. Money, in turn, is of value because of the market-traded goods it allows the individual to consume. Money spent on health interventions is money that cannot be spent on other items. This, of course, is the basic tradeoff that CUA-I analyzes.

From this perspective, an individual's overall welfare can be represented notationally as follows:

$$(6-3) \quad U(.) = U(H,C),$$

where  $H$  is health and  $C$  is the level of material consumption of the individual on (nonhealth) market traded goods.  $U(.)$  stands for overall welfare or utility and the functional notation  $U(H,C)$  simply means that an individual's overall welfare is a function of  $H$  and  $C$  in some unspecified manner. Equations such as (6-3) are often termed "multiattribute utility" specifications. This just means that

overall utility,  $U(.)$ , is a function of more than one attribute. In this case, overall utility is a function of two attributes: health and material consumption.

In order for CUA-I to be firmly linked to VNM-utility theory, little more needs to be assumed about the exact nature of  $U(H,C)$  other than that as health increases, utility increases and that the same is true for material consumption. Technically, however, there is one additional requirement. For CUA-I to be fully consistent with rational choice principles as set forth in utility theory, it must be possible to write an individual's overall utility in the following way:

$$(6-4) \quad U(H,C) = U_1(H) + U_2(C).$$

The assumption that  $U(.)$  can be written in this way is called an *additive separability assumption*. It means that overall utility can be determined by taking some function of health state and some function of consumption and adding them together.

The prior discussion implicitly made this assumption; so seemingly does every CUA of which we are aware. In many instances, the additive separability assumption is an empirically reasonable representation of individual preferences, at least to a first degree of approximation.<sup>18</sup> In essence, the assumption means that the impact of a change in health on overall utility can be assessed independently of the income of the individual (assuming no links between health and consumption other than the direct costs associated with the health care intervention itself). It means, for instance, that an abscessed tooth hurts the same whether one is poor or rich.

## 7. CAN UTILITY BE ATTACHED TO LIFE AND DEATH?

So far we have argued that VNM utility is a compelling way to measure health status for CUA. The argument may appear reasonable as far as it goes, but questions may well arise about its generalizability. In particular, there may be qualms about attaching "utility" values to health outcomes such as death. What does it mean to attach utility values to life and death? How can such values conceivably be used to compare the "value" of life and the "disutility" of death? These questions are similar to those that are often advanced in CBA concerning the "economic value of life." These concerns are not as problematic as may first appear.

Consider a simple example. Suppose an individual faces the risk of a fatal heart attack during the coming year which, for the sake of argument, would result in immediate and painless death. The probability of this occurrence is  $p_1$ . There is a costless health intervention, a drug that is free, which lowers this probability to  $p_2$ . The health intervention results in various side effects, however—lethargy, mild depression, and so forth. Also suppose, for simplicity, that taking the drug during the coming year does not affect the individual's health prospects one way or another after the year is over.

---

<sup>18</sup>See Fischer, Gregory W., "Utility Models for Multiple Objective Decisions: Do They Accurately Present Human Preferences?", *Decision Sciences*, 10, 1979, 451–79. This additive separability does, however, rule out so-called "state dependence" between health and consumption. A small but important literature has emphasized such potential state dependence. See, for instance, Cook, Phillip P. and Daniel J. Graham, "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities," *Quarterly Journal of Economics*, 91, February 1977, 143–56; and, Smith, V. Kerry and William Desvousges, "An Empirical Analysis of the Economic Value of Risk Changes," Vanderbilt University, Department of Economics and Business Administration, Working paper 85-W24, 1985.

Following the prescribed approach to analyzing whether the drug treatment is desirable, the three relevant health states in this situation for the upcoming year can be listed: life with no side effects; life with side effects; and death. Each is assigned a VNM utility. Life, the best of the outcomes, can be assigned a utility,  $U(L)$ , of 1 and death, the worst of the outcomes can be assigned a utility,  $U(D)$ , of 0. Life with side effects has a utility  $U(LS)$ . It can be determined based on a lottery with probability  $p$  of life and probability  $(1-p)$  of death, where the threshold value for  $p$  becomes  $U(LS)$ .

The expected utility theorem indicates that the drug treatment is desirable if the expected utility of drug treatment ( $p_2 \times 0 + [(1-p_2) \times U(LS)]$ ), or  $(1-p_2)U(LS)$  is greater than the expected utility of not having the drug treatment, ( $p_1 \times 0 + [(1-p_1) \times 1]$ ), or  $(1-p_1)$ , where as defined above  $p_1$  is the probability of a heart attack during the coming year if the person does not take the drug, and  $p_2$  is the probability of a heart attack during the coming year if the person does take the drug. So, the drug treatment is preferred if  $U(LS)$  is greater than  $(1-p_1)/(1-p_2)$ . Consistent with common sense, the larger the value of  $U(LS)$ , and therefore the less important the side effects, the more attractive the treatment. Similarly, the smaller  $p_2$  is relative to  $p_1$ , and therefore the more effective the drug in reducing the chance of heart attack, the more attractive the treatment.

But, how can assessments involving life and death be justified? For example, if  $U(LS)$  is .5, how can it be said that in comparison to living without side effects, dying is twice as bad as living with side effects? What do such statements mean? In reality, the individual is not asked in utility theory to assign a "utility" to life or a "utility" to death. This impression is one of the by-products of the perhaps unfortunate choice to call the valuation function in VNM analysis a "utility." Instead, the individual is asked to indicate the increased chance of death that he is willing to bear in order not to experience the given side effects.

Objections might be raised about how any such trade off can be considered. It may be felt that no level of side effects could justify accepting even an infinitesimal increase in the chance of death. This is an acceptable response. If an individual feels this way, he would attach a utility value to Health State LS that is very near or even exactly equal to 1 and would reach the conclusion that the drug treatment is desirable no matter how pronounced the side effects or how small a decrease in the probability of heart attack they imply.

While an individual is free to make this judgement, as the side effects become more severe and the decreased probability of death sufficiently small, most people would probably make the decision to bear the increased risk and not take the drug treatment. Indeed, anyone who drives a car or takes a bus or subway for the "convenience" of going to work has implicitly made such a tradeoff between health and other considerations (assuming one is safer staying home under the covers than venturing into rush hour traffic).

Note also that in this case the individual is comparing the disutility of particular side effects with changes in the probability of death, not contemplating the certainty of his own demise. This way of looking at calibrating life and death in public policy analysis dates back to Schelling (1968). His work shifted the focus of debate from how much a given life is "worth," which seems almost a meaningless question, to tradeoffs individuals are willing to make in order to accept or avoid increased risks. While Schelling's work has been seen as oriented primarily to CBA, its point is equally valid here.

## **8. IMPLICATIONS OF VNM-UTILITY THEORY FOR HEALTH-STATUS MEASURES**

The fact that VNM utility is the correct basis for measuring health in CUA-I has two important implications for the measurement of health states in CUA. The first concerns whether health state measures in CUA should be disease specific. The second concerns whether clinical measures of health outcomes can generally be used in CUA. The short answer to both of these questions is "no."

VNM utility is completely general and is not tied to any specific illness. Fixing a best and worst health outcome, VNM utility allows health states to be measured regardless of the nature of the disease or illness involved. Moreover, if one adopts a convention in which two outcomes are common regardless of the specific illnesses examined (e.g., the best outcome as full health and the worst as death, respectively), VNM-utility allows health states to be measured along a common scale regardless of the nature of the disease or illness involved. This allows, in principle, comparison of the cost effectiveness of health interventions that affect any type of illness. This, indeed, is the common and recommended approach, with full and complete health being given a value of 1 and death a value of 0.<sup>19</sup>

VNM utility's ability to allow health state measurements in CUA to be general and not disease specific is an attractive byproduct of the CUA framework. As argued previously, for CUA to be applicable across many types of health choices, and particularly choices such as between treatment and prevention interventions, it is important that the measure of health in CUA be as broad as possible and be capable of capturing as many different types of health outcomes as possible on a single scale. As we have seen, VNM utility is well designed for this purpose, even to the point of being able to incorporate morbidity and mortality concerns on a common scale.

Use of VNM utility to measure health also rules out certain other health measures for use in CUA. Most clinical measures of health outcomes, for instance, will not be directly applicable to CUA-I. For example, in dealing with health interventions designed to combat high blood pressure, systolic blood pressure, measured in terms of mm Hg, might be a natural clinical measure. It would not in general be valid in CUA, though. This is because a high blood pressure reading of 30 mm hg. over the target may not translate into twice the risk of stroke or heart attack as a reading of 15 mm hg. over the target, and thus would not be viewed as twice as bad, using VNM utility.

In a review of measurement of health status utilities for economic analyses, Torrance (1986) dismisses virtually all clinical health status measures to be found in the literature. While such measures may have clinical validity, Torrance views their determination of magnitudes as "essentially arbitrary" for purposes of CUA. Without wishing to indicate that these measures are not important for other purposes, in their current configuration they are generally not useful for CUA in health care. Work underway to map utilities to psychometric health status instruments holds promise for revision of this perspective.

## **9. RELAXING THE BUDGET-CONSTRAINT ASSUMPTION**

While the cited formulation provides a solid linkage between CUA and VNM-utility theory, at least one aspect of the perspective advanced may seem artificial. This is the so-called budget constraint for

---

<sup>19</sup>This does not rule out health states that are worse than death. Such states can assume negative values, with the results calibrated to the "1—full health/0—death" scale, and all of the above analysis remaining perfectly valid.

health care expenditures. How realistic or general is it to view the individual as facing an externally imposed budget constraint for health? It is true that his total spending may be constrained by his wealth, but there is not generally a prespecified budget constraint for particular categories of expenditure.

Fortunately, the budget constraint is primarily a didactic device, and it can be relaxed without destroying the logic of CUA-I. To see this, suppose that instead of an individual having some preordained amount of money to spend on health care there is a certain amount of total resources—interpreted as income or wealth—to allocate across a wide range of activities or spending categories: health care, food, shelter, recreation, and so forth.

Consider any given allocation of income among these spending categories. Suppose now that an additional dollar were to drop from the sky to be devoted to a given function. This would allow for additional spending in that spending category and would lead to some increased level of VNM utility. Define this increase in utility as the incremental (expected) utility from the marginal dollar spent on the given spending category. Let  $Z_i$  be the incremental utility from the marginal dollar spent on spending category  $i$ .<sup>20</sup>

For an individual to allocate resources across different categories optimally (i.e., so as to maximize overall level of satisfaction), this expected utility for the marginal dollar,  $Z_i$ , should be the same across each category. This is easily seen. Suppose the marginal dollar spent on food provides .6 expected utility units while the marginal dollar spent on health provides .7 expected utility units. Then by moving a dollar from food to health the individual loses .6 units of expected utility but gains .7 expected utility units, for a net increase of .1 expected utility units. Because his goal is to maximize expected utility, such a reallocation is desirable. Whenever there are unequal expected utility returns from the marginal dollar across expenditure categories, such a reallocation can be done. Overall utility is maximized only when such reallocations are not possible, i.e., only when the expected utility return from the marginal dollar in each category is equal. With resources optimally allocated across expenditure categories in this fashion, let  $Z$  be the expected utility that can be reaped from a marginal dollar devoted to any expenditure category.

Now suppose that a new or previously unconsidered health intervention comes to light. Should the individual devote resources to it? The decision can be determined solely by comparing the ratio of the costs of the intervention to the expected increase in VNM utility from the intervention, i.e., by using the cost-effectiveness ratio. If this cost-effectiveness ratio is below  $1/Z$  then the additional expected utility per dollar devoted to this new health intervention is greater than that associated with the marginal dollar devoted to other expenditure areas. That is, the individual will get more expected utility from a dollar devoted to this new health intervention than he or she will lose by not being able to devote it to some other purpose. Thus, if the cost-effectiveness ratio of this new health intervention is below  $1/Z$ , then the intervention should be undertaken and is considered cost effective. By similar reasoning, if the cost-effectiveness ratio is above  $1/Z$  then the intervention should not be undertaken.

---

<sup>20</sup>A natural question arises in terms of the scaling of  $Z_i$ . As discussed previously, health is typically assumed to be additively separable from the other components of overall utility. Under this assumption, it makes sense to talk about the utility of a given health state without having to specify the concurrent level of other attributes. Consistent with our derivation of VNM-utility weights for health states, the best health outcome can be specified and given a utility value of 1 and the worse health outcome can be specified and given a utility value of 0. These utility values then can be used as the basis for scaling all of the  $Z_i$ .

But, this use of  $1/Z$  is identical to the role played by  $R_{\text{Threshold}}$  in the standard mechanics of CUA!  $R_{\text{Threshold}}$  can therefore be defined as follows:

$$(6-5) \quad R_{\text{Threshold}} = 1/Z \quad (= 1/[\partial U/\partial C]).$$

With this definition, the principles of CUA can be used to determine whether to adopt given health care interventions, even without there being any prespecified budget constraint for health care spending.  $R_{\text{Threshold}}$  can thus be viewed not as coming from some arbitrarily imposed budget constraint but from an individual's choice of how to allocate resources optimally among all possible uses, health care being only one of them.

## 10. ADDITIONAL NOTES ON RELAXING THE BUDGET-CONSTRAINT ASSUMPTION

There are several additional points that deserve mention. First, the above analysis assumes that the health care intervention being considered is reasonably small in cost compared to the overall income or wealth of the individual. If the cost to the individual of a given health intervention is very large, then devoting funds to that health care intervention will deprive other activities of so much funding that there will be a new, higher value of  $Z_i$  for these other categories. If this occurs, the value of  $R_{\text{Threshold}}$  for evaluating the new health intervention will decline, meaning the health intervention's cost-effectiveness ratio has to meet a somewhat stiffer criterion for it to be judged cost effective. This complication is similar to the "lumpiness" problem mentioned in Chapter 2.

Second, note that  $Z$  itself is likely to be a function of the level of overall wealth of an individual. It is assumed usually that a VNM-utility function is such that increments in utility from a given increase in wealth decrease as an individual's wealth increases. This assumption of diminishing marginal utility means that, while an individual would like to be given an additional \$1,000 regardless of whether he is rich or poor, he gets more utility from this \$1,000 when he is poor than when he is rich. For our purposes, this means that the more wealth an individual has, the lower is  $Z$ , the incremental utility from another dollar of consumption.

This has implications for  $R_{\text{Threshold}}$ . In particular, the wealthier an individual is the greater  $R_{\text{Threshold}}$ . This makes perfect sense. Health interventions that previously cost too much to be judged cost effective when the individual was poorer now are cost effective. Being wealthier, the individual is more willing to devote money to a health intervention even if its health return is smaller than he previously would have judged necessary in order to justify its cost.

Finally,  $R_{\text{Threshold}}$  also can be viewed as the "imputed monetary value" the individual places on an expected unit of health. To see this, note that under the decision-making scheme just derived, the individual will choose to adopt a given health intervention,  $HI_i$ , if  $C_i/[\Sigma p_{i,n}U(\Delta h_{i,n})]$ , its cost-effectiveness ratio, is less than  $R_{\text{Threshold}}$ . But this will be the case if and only if  $[\Sigma p_{i,n}U(\Delta h_{i,n})] \times R_{\text{Threshold}} - C_i > 0$ . That is, a health intervention will be adopted only if the imputed monetary value associated with the expected utility of the health change from the intervention exceeds the cost of the intervention.



## CHAPTER 7

### CUA-II: THE SOCIAL PERSPECTIVE

#### 1. INTRODUCTION

Chapter 6 considered CUA from the perspective of an individual making personal health care resource allocation decisions. Can this same approach be extended to using CUA as a guide to resource allocation for governmental health policy?

One way to move from an individual to a societal or government perspective is to exploit the result obtained in Part I that CUA, and its associated mechanics based on a cost-effectiveness ratio, is mathematically equivalent to maximizing the aggregate health benefits from a given expenditure of health care resources or minimizing the cost of achieving any given level of health benefit. As Weinstein and Stason (1977) write,

the underlying premise of cost-effectiveness analysis in health problems is that, for any given level of resources available, society (or the decision making jurisdiction involved) wishes to maximize the total aggregate health benefits conferred. Alternatively, for a given health-benefit goal, the objective is to minimize the cost of achieving it.<sup>21</sup>

This alone provides some justification for the use of CUA as a societal guide to health care allocation. It may be that a government agency is charged with maximizing aggregate health benefits subject to its given budget constraint. Then, CUA naturally becomes the analytic tool of choice.

However, for many purposes this leaves open too many fundamental questions, the answers to which seem critical to understanding the strengths and limitations of CUA in health care. It fails to address *why* the total aggregate health benefits should be maximized or *why* health outcomes should be measured in terms of VNM-utility. As such, it is important once again to consider the normative underpinnings of CUA in more detail.

#### 2. A CONSTITUTIONAL CONVENTION ON HEALTH CARE

Consider members of society getting together in a sort of constitutional convention, called together to come up with a set of procedures that society should employ in allocating health care resources. Some choice mechanism could be imposed on the convention, such as majority rule. It may be felt that "socially appropriate" decisions concerning health care resource allocation can then be judged in terms of whatever this convention agrees upon. What sort of rules or procedures will such a convention adopt?

There has, in fact, been much work within "public choice theory" about the types of social choices that would result in this type of setting. Public choice theory suggests that, under certain circumstances, majority rule leads to a decision which is in accord with the preferences of the so-called "median voter." The median voter is the voter who is right in the middle, with an equal number on the left and the right. It was very quickly realized in the public choice literature,

---

<sup>21</sup>Milton C. Weinstein and William B. Stason, *Foundations of Cost-effectiveness Analysis for Health and Medical Practices*, *New England Journal of Medicine*, 296(13):716-21, 1977, p. 717.

however, that this median voter theorem holds only in special situations. Without going into the details here, it can be asserted that these conditions are unlikely to hold generally for individuals trying to make decisions about societal health-care allocation rules.

Moreover, when the conditions required by the median voter theory do not hold, it can be shown that it is likely that such a constitutional convention would not reach *any* consensus about the rules to endorse. For any rule that might be considered, there is generally some other rule that would beat it in a fair, majority rule election. The result would be an indeterminate cycle in which Health Policy Rule B beats Health Policy C in an election, Health Policy Rule A beats Health Policy Rule B in an election, but Health Policy Rule C beats Health Policy Rule A, and so forth. The public choice literature thus suggests that this sort of constitutional convention may not lead to any set of stable, predictable outcomes.

This conclusion is not necessarily as disheartening as it may first appear, however, in thinking about socially desirable health care policy. This is because of the self-interest that would guide voting at such a constitutional convention. Each individual in the convention would try to promote his own self-interest, with the person suffering from heart disease arguing that all available funds should go to heart-related illnesses and none for mental health care while the person suffering from depression would argue just the opposite. This pursuit of self-interest would seem to preclude a just or fair health care policy.

Without defining "fairness" or "justice" with precision here, it would seem that the determination of a fair or just policy requires at the least that decision makers distance themselves from their own self-interest. It is with this idea in mind that many ethicists and social scientists interested in normative analysis have suggested a modification to the above constitutional convention framework. This modification carries substantial appeal and also may resolve the indeterminateness about which social choice rule will be adopted.

Suppose that the constitutional convention was somehow set up so that the individuals at the convention did not know which policies would benefit them individually as compared to other people. Suppose, for instance, decisions had to be made about health policy for a series of illnesses each member was equally likely to experience in the future. Or, in a similar spirit, suppose the members of the constitutional convention are called upon to make health policy rules not for themselves but for future generations. In such circumstance, the individuals at the convention can be described as making their decisions under a "veil of uncertainty."<sup>22</sup> They would not know the degree to which they individually will be affected by a given policy.

The notion that "fair" or "just" decisions can be thought of as those which would be made by individuals behind a "veil of uncertainty," in which self-interest is muted or eliminated, has been the

---

<sup>22</sup>The term "veil of uncertainty," as used by Buchanan and others, was undoubtedly chosen to reflect the similarity with the influential work of Rawls. Rawls, a philosopher who has greatly influenced modern thought about the notions of justice and fairness, coined the terminology "veil of ignorance" in a context of a particular type of thought experiment. In this thought experiment one is asked to judge among alternative ways by which society may function by imagining that one will be a member of society but that one is not yet told which member of society one will be. The choice made behind this "veil of ignorance" about one's own position in society becomes Rawls's basis for determining justice and fairness. Despite many thematic linkages with this Rawlsian notion, the approach and conclusions of the Buchanan and Tullock framework differs in subtle but important ways from Rawls, and thus the perhaps purposeful distinction in the terms used.

basis for many influential works in the social sciences. This can be formalized in the following way, borrowing from Buchanan and Tullock (1962) and others. Let society consist of  $N$  members. These  $N$  members are brought together to set forth constitutional rules for societal health policy. At this constitutional convention, individuals are in an initial state in which they are "yet to be born." In particular, they do not know which members of society they will be. They are able to project the impact of different health policies to the welfare and quality of life of different members of society. They do not, however, know which individuals they will be in this society and therefore do not know whether they will be sick or well, which illnesses they will have, and so forth.

After the constitutional convention has decided on a set of constitutional provisions for national health policy, individuals will then enter a lottery in which they are assigned to become specific members of society. Each individual has an equal chance of  $1/N$  to be any specific member of society. Each individual knows this when deciding on constitutional provisions. As described by Buchanan and Vanberg (1990),

The participant who is to vote in the constitutional convention...remains, by presumption, ignorant or uncertain in the Rawlsian sense. Each and every participant remains unable to identify his own interest in relation to the working of the choice alternatives in periods after the constitutional choice has been made by the group. This sort of Rawlsian constitutional ignorance may facilitate agreement because it offers a motivation for participants to make choices in terms of generalizable rather than particularized criteria.<sup>23</sup>

Several points should be made about this general perspective. First, individuals at the constitutional convention are presumed to remain "self-interested." Each individual wishes to make himself as well off as possible through the choice of social policies at the convention, but he does not know which individual in society he will be. Thus, any "societal interests" are forged solely by the veil of uncertainty, rather than by any abstract notions of societal well-being or altruism.

Second, while the discussion has been oriented toward justifying this constitutional perspective as one promoting fair and just policies, this is not to indicate that traditional economic concerns with efficiency cannot also be taken into account. Indeed, one of the strong appeals of the constitutional perspective is that it is fully consistent with the Pareto principle.<sup>24</sup> To see this, imagine a choice between two provisions before the convention, Provision A and Provision B, where Provision B leaves everyone better off than Provision A. Under the framework set out above, Provision B would be preferred unanimously to Provision A by everyone at the constitutional convention. While members of the constitutional convention would not know which individuals in society they are to become, they do know that no matter who they become, they will be better off under Provision B than Provision A. Thus, the Pareto principle, which is the buttress for the notion of economic efficiency, remains a compelling ethical principle from a constitutional perspective.

---

<sup>23</sup>James Buchanan and Viktor Vanberg, "A Theory of Leadership and Deference in Constitutional Construction," *Journal of Constitutional Economics*, forthcoming.

<sup>24</sup>The Pareto principle is the linchpin of economic welfare analysis. It states that an action that makes some individuals better off without leaving any worse off is socially desirable. This principle serves as the basis for the definition of economic efficiency. A situation is economically efficient if it is not possible to take some action that would make some individuals better off and leave no one worse off. (Vilfredo Pareto, *Trattato di Sociologia Generale*).

The basic ethical framework of a convention behind a veil of uncertainty is closely related to Schelling's *ex ante* perspective on valuing health states, in which individuals make tradeoffs in order to accept or avoid increased risks (discussed briefly in Chapter 6 in the context of the "utility of a life"). This framework also is related to what Gibbard (1983) labels "*ex ante* Pareto optimality." While an admittedly abstract way to think about setting health policy guidelines, many ethicists and normative social scientists find this sort of thought experiment a powerful way to address issues of social choice. They feel that the kind of rules that would emerge from such a constitutional convention to be ones which separate self-interest from societal interest and which adhere to notions of fairness and justice.

### 3. LINKING CONSTITUTIONAL CHOICE TO VNM UTILITY

To link constitutional choice to VNM utility and CUA, we consider a simplified world which helps bring out the key relationships. For present purposes, assume the following:

1. Health care resources, measured in dollars, have fallen from the sky in a predetermined amount per unit time. No other money may be spent on health care.
2. Health care resources improve health but only when an individual needs them. Whether a particular individual benefits if provided a given type of health care is something which is beyond the control of the individual himself.
3. Individuals experience illness similarly. They rank the health states they might experience in the same order. Fixing the best and worst health outcome with utility 1 and 0, respectively, they assign the same VNM-utility values to the remaining states.
4. A health intervention affects only health and requires only health care resources.

Within this world, consider the nature of constitutional choice for health policy from behind a veil of uncertainty. To begin with a simple example, suppose the population consists of only two people. It is known with certainty that both individuals in the upcoming year will catch illness X. It costs \$10 to treat illness X. If an individual has illness X and health care resources are provided, the individual is immediately made completely well. It is also known with certainty that one of the two individuals will catch illness Y during the coming year. Neither individual knows if he will be the person that catches illness Y. It costs \$20 to treat illness Y. If an individual has illness Y and health care resources are provided, the individual is immediately made completely well. There is a total of \$20 to be devoted to health care during the next year. The two individuals must decide before the year starts whether to devote the money to treating illness X or illness Y.

What rule will the constitutional convention choose? In this example, there are four health states that each individual can consider. State 1 is being well. State 2 is having illness X. State 3 is having illness Y. State 4 is having illness X and Y. State 1, being well, is the best state. State 4, having both illnesses, is the worst state. If State 2, having illness X, is worse than State 3, having illness Y, the convention will choose to have the resources devoted to treating illness X. Even the person unlucky enough to get illness Y would rather have the resources devoted to treating illness X.

More interesting and difficult is the question of the constitutional choice when State 3 is worse than State 2. Now the individuals face a trade-off. They must choose between applying the resources to treat illness Y for one of them while leaving both with illness X or treating illness X while each takes

the chance of not being the one to get illness Y. Consider one of the individuals. He or she views the choice as follows: if resources are devoted to combat illness X then, since one of us will catch Y and I don't know if it will be me or not, there is a .5 probability I will be well and a .5 probability I will have illness Y; if resources are devoted to combat illness Y then I will have illness X for sure, since we will both catch illness X.

This choice, however, is exactly the situation modeled in Chapter 6 in terms of VNM utility. The individual will consider the utility of illness X,  $U(X)$ , and the utility of illness Y,  $U(Y)$ , where the best outcome, being well, has a utility of 1 and the worse outcome, having both illnesses, has a utility of 0. If health care resources are devoted to illness Y, then the individual has an expected utility of  $U(X)$ , because he knows with certainty he will have illness X. If health care resources are devoted to illness X, then the individual has an expected utility of  $(.5 \times 1) + (.5 \times U(Y))$ , since there is a .5 chance he will have neither illness (utility equals 1) and a .5 chance he will have illness Y (utility of  $U(Y)$ ).

The individual will prefer that health care resources go to illness Y if the expected utility of that option,  $U(X)$ , is greater than the expected utility of devoting health care resources to illness X,  $.5 + U(Y)$ . So, given the probabilities we have specified, the individual will prefer health care resources go to illness Y if  $U(X) > .5 + .5U(Y)$ . This is sensible; the greater  $U(X)$  relative to  $U(Y)$ , the less benefit from treating illness X relative to illness Y and the more attractive is the option of devoting resources to illness Y. If  $U(X) < .5 + .5U(Y)$  then he will prefer that health care resources go to illness X. If  $U(X) = .5 + .5U(Y)$ , then he is indifferent between the two allocation options.

Moreover, both individuals reason exactly the same because, as assumed above, they experience illnesses X and Y comparably in terms of the utility or disutility. That is, they assign the same values for  $U(X)$  and  $U(Y)$ , measured relative to the options of being well and having both illnesses. So the constitutional convention will adopt the rule to devote health care resources to illness Y and not to X if  $U(X) > .5 + .5U(Y)$ . If  $U(X) < .5 + .5U(Y)$ , the convention will adopt the rule to devote health care resources to illness X.

Compare this constitutional choice to what would occur if the decision about health care resources were made not before, but after it was known who is to catch disease Y. In that case, the individual who was not to catch disease Y would, in his own self-interest, want resources devoted to illness X. If Y is worse than X, as we are supposing, the individual with illness Y will want the resources devoted to illness Y. So, compared to the unanimous agreement on health policy that results from a constitutional convention, making choices after health states are known guarantees disagreement among the individuals; each, pursuing his/her own self-interest, wants health care resources devoted in such a way that he is made better off. This constitutional choice thought experiment thus allows the issue of health care resource allocation to be framed in such a way that self-interest of the parties is removed and allows an "impartial" determination of a "fair" or "just" allocation.

#### 4. LINKING CONSTITUTIONAL CHOICE TO CUA

The example from the prior subsection links the choice that individuals would make about health care policy in a constitutional setting to VNM utility. What does this have to do with CUA-II? To see the link, consider the types of choices that would be made using CUA, where CUA evaluates a health intervention based on the ratio of the cost of a health intervention to the change in aggregate health status that results from the intervention, and where aggregate health is the sum of the VNM utilities of the individuals for the health states they experience.

Consider a set of health interventions,  $HI_1, HI_2, \dots, HI_k$ . Let each health intervention,  $HI_i$ , have a cost of  $C_i$ . Let there be  $N$  individuals affected by the health intervention. Let Health Intervention  $i$  (deterministically) have health impact  $\Delta h_{i,n}$  on individual  $n$ . Then, the resulting cost-effectiveness ratio for Health Intervention  $i$  is:

$$(7-1) \quad C_i / [\sum_n U(\Delta h_{i,n})].$$

Or, if Health Intervention  $i$  has a stochastic health effect on individual  $n$ , then this formula becomes:

$$(7-2) \quad C_i / [\sum_{n,j} \{p_{n,i,j} U(\Delta h_{i,n,j})\}],$$

where  $p_{n,i,j}$  is the probability that individual  $n$  will experience health impact  $\Delta h_{i,n,j}$  from Health Intervention  $i$ .

Let us apply this approach to the example of the previous section. The default alternative (e.g., the status quo) from that example can be viewed as devoting nothing to treating either illness X or Y. In this case, one person will have illness X and the other will have illness X and Y. Having both illnesses X and Y was the worst of the outcomes considered and was assigned a utility value of 0. The person with both illnesses would thus have a utility of 0. The person with illness X was assigned a utility of  $U(X)$ . The aggregate health measure when there is no allocation of resources would therefore be the sum of the individual health measures under this scenario, or 0 plus  $U(X)$ , or simply  $U(X)$ .

The two alternatives to doing nothing, each costing \$20, are to devote resources to treating illness X or to treating illness Y. Consider initially the alternative of devoting the \$20 to treating illness X. This will leave one individual with illness Y and the other individual completely well. So, one individual will have utility  $U(Y)$  and the other will have utility 1. The aggregate health measure is then  $1 + U(Y)$ .

The change in aggregate health between this intervention and the default is  $1 + U(Y) - U(X)$ . The CUA ratio for the first intervention is then  $\$20 / (1 + U(Y) - U(X))$ . This indicates the dollars per health utility unit associated with treating illness X.

The other alternative to doing nothing is to devote the \$20 of health care resources to treating illness Y. This will leave both individuals with illness X. The aggregate health measure will then be  $2U(X)$ . The change in aggregate health between this intervention and the default is  $2U(X) - U(X)$ , or  $U(X)$ . The CUA ratio for the second intervention is then  $\$20 / (U(X))$ . This indicates the dollars per health utility unit associated with treating illness Y.

The CUA ratios for both interventions are positive, indicating that for either intervention the (positive) health care resources lead to increased aggregate health. Since they both involve the same cost, CUA indicates that the more desirable of them is the one with the lower CUA ratio. Thus, from a CUA perspective, Intervention 1, treating illness X, is preferred if  $\$20 / (1 + U(Y) - U(X)) < \$20 / U(X)$ , or if  $U(X) < .5 + .5U(Y)$ . So, based on CUA, resources should be devoted to treating illness Y if  $U(X) < .5 + .5U(Y)$ . If  $U(X) = .5 + .5U(Y)$ , then CUA indicates that both treatments are equally desirable.

But these are exactly the conclusions from the earlier analysis based on the constitutional convention thought experiment, as derived in the prior section! Moreover, there is nothing "rigged" about this

example. No matter what numbers had been chosen or in what other ways this example could have been specified, an exact correspondence between the unanimous recommendation of the constitutional convention and the recommendations using CUA ratios would have been reached, where health outcomes are measured using VNM utility and where aggregate health outcomes are measured as the sum of the individual VNM utilities. Thus, we reach the following conclusion: the unanimous constitutional choice as to the allocation of health care resources amounts to the computation of CUA ratios, where individual health outcomes are measured in terms of VNM utility and aggregate health outcomes are measured as the sum of the individual measures.

## **5. MERGING INTERPERSONAL AND INTRAPERSONAL COMPARISONS**

This basic normative framework allows the two key interpretive questions for CUA raised in the Introduction to Part II to be addressed. The first interpretive question, involving assignment of magnitudes to health status changes for an individual, was addressed in the discussion of CUA-I and the rationale remains identical here. The second interpretive question involves comparisons across individuals: why is a change of three units in health state for one individual "better" than an improvement in one unit in health state for two individuals? In short, how can one derive an aggregate health measure that can merge considerations about the degree to which a given individual's health is improved and about the number of individuals helped?

Justifying CUA from a constitutional perspective provides a clear response to this question. From this perspective, an improvement of three units in one person is "better" than an improvement of one unit for two individuals because of the following reasoning that someone behind the veil of uncertainty would use. Suppose there are two individuals in society and you don't know which of them you will be. You have a .5 chance of being either of them. Prior to learning which individual you will be you have the choice of improving one person's health state by three units or both persons health state by one unit. Given the definition of VNM utility you choose the former simply because your expected utility after learning who you will become is larger that way. It is true, in a sense, that you increase your risk by this choice. By improving both individuals by one unit, you are guaranteed a better position than if you improve one person by three units but are unlucky enough to be assigned the person who is not improved. Nonetheless, the logic of expected utility maximization indicates that the improvement by three units in the one person is sufficiently large to make you willing to assume this risk.

Thus, the answer to the second interpretive question follows directly from the conclusion that from behind the veil of uncertainty an individual will wish to maximize his expected utility from the health state that will be assigned after the veil of uncertainty is removed. Because each individual views himself or herself from this perspective as having an equal chance of being assigned to any of the health states in the population, it is immaterial whether a given increase in a unit of VNM-utility is given to one person or divided among several. From behind the veil of uncertainty any division of this health status unit results in the same level of overall desirability.

## **6. $R_{\text{THRESHOLD}}$ AND CUA-II**

The analysis of the prior sections indicates that CUA-II can be performed by applying the mechanics presented in Part I and by using cost-effectiveness ratios presented in Equations 7-1 and 7-2. The result will be decisions about health care allocation that are consistent with the normative framework that has been proposed. In the analysis of the prior sections, it was assumed that a set amount of money was available for health care (\$20 in the example used throughout the chapter). This leaves

unanswered, however, the question of how much money should be spent on government health care programs in the first place.

This is a difficult issue and not completely separable from even more basic questions about what types of health programs should be funded by the government and why. While the issue is not resolved here, the framework presented above does allow for some insight into the question. First, it can be noted that any fixed amount of money for government health care programs implicitly defines a value for  $R_{\text{Threshold}}$ . It does this in exactly the fashion described in Part I. For an available pool of health care resources and a set of mutually compatible health intervention, the health interventions can be listed in order from those with the lowest cost-effectiveness ratio to those with the highest. The list can then be gone down in order until the available funds are used up. The value of the cost-effectiveness ratio of this last approved program then is defined as  $R_{\text{Threshold}}$ .

Moreover, as indicated in Part I, important conclusions about health care allocation can often be provided without determining a specific value of  $R_{\text{Threshold}}$ . For instance, substantial progress can be made in narrowing the list of mutually exclusive programs without an explicit value for  $R_{\text{Threshold}}$ . Additionally, if Health Intervention  $i$  has a higher cost-effectiveness ratio than that of Health Intervention  $j$ , then statements like the following can be made: "If it is judged cost effective to spend resources on Health Intervention  $i$  then it is even more cost effective to spend resources on Health Intervention  $j$ . Moreover, if there is not enough funding available for both, then Health Intervention  $j$  should be preferred from the perspective of CUA." Similarly, it can be stated that "if it is judged cost ineffective to spend resources on Health Intervention  $j$  then it is also cost ineffective to spend resources on Health Intervention  $i$ ."

The current state of the art in CUA goes little further than this and for most purposes it is enough. Typically, a cost-effectiveness ratio for a health intervention is computed in dollars per quality-adjusted life-year (\$/QALY). This \$/QALY is then compared with that of other health interventions to gauge social desirability. As a very rough rule of thumb, some authors suggest that \$50,000 dollars per QALY is reasonable for the threshold in many policy settings<sup>25</sup>. We would not feel very comfortable with this as any sort of hard and fast norm, but a health intervention that costs under \$10,000/QALY usually is considered a very cost-effective use of health care resources while one that costs over \$100,000/QALY is seldom seen as a noteworthy return to aggregate health from societal health care resources.

Can one be more specific about the proper value of  $R_{\text{Threshold}}$ ? In addressing this issue, two contrasting positions can be identified. From one position, the amount of funding available for government health care programs is determined by the political system. The political system's judgement on this issue may be wise or not, but it is not a proper or useful task for a health policy analyst to question this judgment. It serves little purpose to apply a value of  $R_{\text{Threshold}}$  that is at extreme variance with that implicit in societal values, as reflected by the level of funding made available for government health care programs.

From another position, there is value, at least as a point of comparison, in considering how the "proper" level of  $R_{\text{Threshold}}$  might be thought about from a normative perspective such as the one advanced here. Suppose a constitutional convention, making its decisions from behind a veil of uncertainty, is charged with determining a value for  $R_{\text{Threshold}}$ . Suffice it to say here that in very

---

<sup>25</sup>Dennis G. Fryback and John R. Thronbury, "The Efficacy of Diagnostic Imaging," *Medical Decision Making* Vol 11, No. 2, Apr-June 1991, p. 89.

simple situations, some insights can be provided on a theory of the endogenous choice of  $R_{\text{Threshold}}$ . However, complications quickly arise. For one thing, the choice of how much to provide in funding for Government health care programs will be influenced by the different ways in which the funds are raised, that is by the choice of taxation schemes. An additional complication arises if health care is provided by the Government but can be augmented by an individual. Further complications arise concerning incentive effects. One incentive effect is "moral hazard" in which individuals may be less careful about their health if they are automatically covered for health care expenses. Another incentive effect arises from the effect of taxes for Government health care programs on labor supply decisions. These all represent concerns that are substantially beyond the current state-of-the-art in CUA and which will not be explored further in this monograph.

## 7. SOME CONCLUDING COMMENTS

The conclusion from a constitutional choice perspective is that, at least under the simplifying conditions given, society should choose to allocate its health care resources such that it maximizes the sum of the utility for members of society, or in notational terms,  $\sum_n U(\Delta h_n)$ , where  $n$  indexes the individuals in society and  $U$  represents an individual's VNM utility. This, in turn, becomes the implicit objective achieved by CUA as we have defined it.

This is the same conclusion proposed by the traditional philosophical school of thought known as "utilitarianism." Thus, the perspective can also be termed a utilitarian perspective. The main difference between our approach and utilitarianism proper is one of motivation. Utilitarianism views this goal as an initial axiom and starting point. We, in contrast, derive it as the conclusion to constitutional choice.

While utilitarians may feel comfortable with the conclusion of the above analysis, if not with all of the details of how the conclusions is reached, not all ethicists or social scientists concerned with normative welfare policy will embrace this perspective. One might immediately contrast this framework with a Rawlsian framework or rights-based frameworks such as proposed by Daniels (1985), adherents to either of which would presumably have little sympathy or support for CUA in health care in anything resembling its current presentation.

There is nothing wrong with these differences of perspective. The point of the exercise here is not necessarily to persuade the reader that the constitutional perspective is in some sense *the* correct perspective on social choice. The point instead is to indicate explicitly the type of normative framework implied by CUA. An individual is free to accept or not accept this framework. We would suggest, however, that if one finds CUA in health care to provide important guidance for the allocation of health care resources, then one is also accepting something very much like the underlying normative framework proposed here. Conversely, if one finds the underlying framework proposed here to be unacceptable, then CUA in health care is likely to be of little persuasiveness as a guide for the allocation of health care resources.

If the normative framework put forth here *is* accepted, then CUA-II becomes a very natural extension of CUA-I and the basic orientation and procedures of CUA-I presented in Chapter 6 remain largely in place. CUA-II continues to use a cost-effectiveness ratio and to apply the mechanics put forth in Part I. Now, instead of individual health benefits, the denominator of the cost-effectiveness ratio is aggregate health benefits. Similarly, aggregate costs replace individual costs in the numerator. This, though, does not represent a fundamental reorientation of the principles put forth in Chapter 6. Costs and health outcomes continue to be measured largely as previously, but now they are extended across

the individuals affected by a health intervention. VNM-utility remains central to the measurement of health outcomes. Thus, CUA-II builds directly on CUA-I.

## CHAPTER 8

### THE RELATIONSHIP OF CUA TO CBA, DA, AND COIA

#### 1. INTRODUCTION

The history and development of cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) in health care is interwoven with the history of cost-of-illness analysis (COIA), cost-benefit analysis (CBA), and decision analysis (DA). The linkages between these analytic methods for evaluating health interventions are often a source of confusion. The goal of this chapter is to explore the relationships among these methodologies and to indicate ways in which they are linked and ways in which they properly remain distinct. In so doing, these other methodologies will occasionally be contrasted not only with CUA but with CEA more generally conceived since many of the debates in the literature pertain to conceptions of CEA as well as the more narrowly defined CUA.

#### 2. COST-OF-ILLNESS ANALYSIS

Health economics began to emerge as a field in its own right only in the late 1950s (e.g., Mushkin, 1958). The first systematic quantitative analyses in the field followed shortly thereafter, in the form of COIAs. The seminal early work is Rice (1966). COIA is not primarily oriented toward the evaluation of given health interventions and thus is not directly comparable with CUA. The evolution of COIA and its measurement procedures, however, shed important light on CUA as well.

COIA measures the costs of an illness in a population, where costs are defined as they are in the national income accounts. That is, a COIA measures the effect of an illness on a country's gross national product (GNP). Loosely speaking, a COIA indicates how much greater the GNP would be if a given illness were eradicated.

Consistent with national income accounting, there are two types of costs associated with an illness. The first is *direct costs*. Direct costs are the costs of the medical services provided in treating or preventing an illness. These consist of medication, the time of health professionals, and so forth. The second type of costs is *indirect costs*. Indirect costs are the effect of an illness on the income of an individual (or on the labor costs of his/her employer). These consist of lost work due to mortality, and absenteeism and reduced productivity due to morbidity. This way of measuring costs is often referred to as the "human capital" approach.

Several sets of controversies arose within the human capital and COIA literatures. One issue was whether, when an individual died prematurely from a given illness, the indirect cost was based on his future income or his future income minus his future consumption. This issue was never resolved although most analyses choose the former approach. The lack of resolution reflects the lack of a clear normative foundation for COIA.

A second issue in the COIA literature stems from the fact that COIAs seem to discriminate against women, homemakers, the unemployed, minorities, and the elderly, all of whom are out of the workforce or earn less money than others.<sup>26</sup> In counteracting this, some more recent COIAs attempt

---

<sup>26</sup>For instance, using a COIA approach, the life of a typical male in his early thirties would be estimated to be "worth" \$430,000 while that of a comparably aged woman in the workforce to be "worth" only \$290,000. See Dorothy P. Rice, Thomas A. Hodgson, and Andrea N. Kopstein, *The Economic Cost of Illness: A Replication and Update*, *Health Care Financing Review*, 8(1), 1985, 61-80.

to assign economic impacts to effects borne by homemakers. For instance, attempts are sometimes made to determine how much it would cost to hire individuals to perform the task of a homemaker and to use this figure as the economic impact of illness on a homemaker.

In general, economic analysis has tended to move away from the human capital approach, and therefore away from COIA, in valuing the effects of illness. The human capital approach was seen as not reflecting accurately the true value of health to an individual, because of the above-mentioned bias against those out of the work force and also because it ignored pain and suffering. In the extreme, the approach leads to absurdities, such as the seeming desirability of a retired person dying from an illness rather than living and accumulating future health costs associated with that illness.

### 3. CUA AND COST-BENEFIT ANALYSIS

#### 3.1. Introduction

CBAs in health care can be traced back almost as far as COIA (e.g., Klarman, 1967). The Office of Technology Assessment (1980) reports an average of six CBA health care articles per year during 1966 through 1969, increasing to over ten per year by the early 1970s. At first CBA studies outnumbered CEAs by 70 percent to 30 percent. By the early 1970s, however, CEAs began to outnumber CBAs. As Emery and Schneiderman (1989) write, "Cost-effectiveness analysis has rapidly become favored over CBA in the health care sector."<sup>27</sup>

Despite the fact that CEA (defined broadly, including CUA as considered here) has come to dominate CBAs in health care in terms of "market share" of empirical analyses, CEA and CUA are often posited as intermediate steps towards CBA. Often, the key difference is seen as an aversion of some sort that CEA or CUA has in attaching dollars to health. This is the position adopted by Office of Technology Assessment (OTA) (1980), for instance, which typically referred to CEA and CBA together as CEA/CBA. Weinstein and Stason (1976) also reflect this view when they write,

...the major disadvantage of the benefit-cost framework is the requirement that human lives and quality of life be valued in dollars. Many decision makers find this point difficult, and do not trust analyses that depend on such valuations. Cost-effectiveness analysis, on the other hand, requires only that health outcomes be expressed in commensurate units.<sup>28</sup>

In a similar vein, Russell et al. (1988) argue,

Indeed, the distinction between cost-effectiveness and cost-benefit analysis is this way of measuring outcomes—cost-benefit goes one step further, often a controversial step, and assigns a dollar value to the health outcomes.<sup>29</sup>

---

<sup>27</sup>Emery and Schneiderman, p. 12.

<sup>28</sup>Weinstein and Stason, p. 717.

<sup>29</sup>Louise Russell, William Taylor, Marianne Fahs, David Rose, Anthony Komaroff, Jeanne Mandelblatt, Theodore Pass, Clyde Schechter, Donald Shepard, and Alan Silver. *Some Examples of the Cost-effectiveness of Preventive Services: Background Paper for the U.S. Preventive Services Task Force*, unpublished, 1989.

Mishan (1988), getting right to the point in his widely used text on cost-benefit analysis, writes,

To be rather rude about it, the analysis of cost-effectiveness can be described as a truncated form of cost-benefit analysis...It should be emphasized, however, that useful as cost-effectiveness analysis may be, a cost-benefit analysis is in principle superior.<sup>30</sup>

From this perspective of CEA as a step to CBA, health is separated from other benefits and costs in CEA simply because assigning a dollar figure to health is often in practice a difficult and controversial component of the analysis. It is therefore best to place this step last. Separating health outcomes and other benefits and costs, as in CEA/CUA, serves to do this. In so doing, CEA/CUA makes this controversial step of assigning a dollar figure to health outcomes more explicitly, prevents it from being hidden among other calculations and procedures, and allows those with different opinions about the proper dollar figure for health outcomes to see more easily the implications of their own views of this issue.

Before turning to the question of whether CUA, as defined here, is indeed an intermediate step to CBA, it is useful to note a shift in the measurement of costs in most CBAs compared to COIA. Beginning in the early 1970s, and in reaction to dissatisfaction with the human capital approach, CBAs in health care began to shift to the concept of "willingness to pay" for the monetary measurement of health outcomes.

Here, too, though, there were concerns. The willingness-to-pay perspective appears to require that one estimate how much one would be willing to pay to save one's life or, conversely, how much one would have to be paid in order to give up life voluntarily. Both questions appear to many to be virtually meaningless and of dubious usefulness in guiding health care resource allocations.

However, a concurrent shift in orientation did much to assuage these concerns. Beginning with Schelling (1968) and formalized further by Mishan (1971) and Jones-Lee (1976) the perspective shifted from asking how much one would pay to save one's life to how much one would pay for a small incremental increase in safety (e.g., from a chance of death in an automobile accident from 20/100,000 during the next year to, say, 18/100,000). This was seen as a sensible question and represented the kind of tradeoff between health and money that people frequently make. The economic value of a life thus became not the value of a given life but the value of a "statistical life."<sup>31</sup> The reader should be able to note quickly the influence of this approach on CUA as well, as discussed in Chapter 6, Section 7, "Can Utility be Attached to Life and Death."

---

<sup>30</sup>E.J. Mishan, *Cost-Benefit Analysis: An Informal Introduction*, Fourth Edition, Unwin Hyman, London, 1988, p. 110 and 113.

<sup>31</sup>Suppose an individual indicated he would be willing to pay \$10 for an increase in safety from an automobile accident from 20/100,000 to 18/100,000. This change in odds means that for every 50,000 people one person who would otherwise have been killed in an auto accident during the next year will not be killed. If each person values this increase in safety by \$10, then one might say as a form of short hand that the economic value of a life is  $50,000 \times \$10$ , or \$500,000.

### 3.2. The Mechanics of CBA

With this introduction to CBA, we can now turn to the basic mechanics of CBA and use them to evaluate whether CUA as defined here can be viewed as an intermediate step to CBA. Consider a health intervention that requires some health care resources and results in some aggregate health benefits. How does CBA go about analyzing such a health intervention? In CBA, government programs are evaluated based on their net (social) benefit, NB, where net benefit is defined as the difference between total benefits, TB, and total costs, TC. Both benefits and costs are measured in terms of the willingness to pay of those affected by them. Total benefits thus are measured in terms of the willingness to pay of individuals for the health benefits received. Total costs are measured in terms of the (social opportunity) cost of the health care resources involved in the health intervention. Both benefits and costs are measured relative to the status quo.

The evaluation criteria proposed by CBA in evaluating health programs can then be summarized as follows:

- In comparing a given health intervention, Health Intervention  $i$ , to the status quo, CBA recommends a health intervention if it has positive net benefit, i.e., if  $NB_i = TB_i - TC_i > 0$ .
- In choosing among mutually compatible health interventions,  $HI_1, HI_2, \dots, HI_k$ , all interventions with positive net benefit should be funded, i.e., all  $HI_i$  for which  $NB_i = TB_i - TC_i > 0$ .
- In choosing among mutually incompatible health interventions, the intervention with the greatest net benefit,  $NB_i$ , should be chosen.

### 3.3. CUA-I as an Intermediate Step to CBA

From the perspective of an individual choosing among health interventions that affect him or her, there is a direct relationship between the procedures of CUA and CBA. This can be seen by taking the framework set forth in Chapter 6 for CUA-I and applying it to the basic mechanics of CUA presented in Part I.

Suppose, as above, that there are several health interventions,  $HI_1, HI_2, \dots, HI_k$ . Each health intervention,  $HI_i$ , is associated with costs,  $c_i$ , to the individual and with a change in the individual's health of  $\Delta h_i$ . (For simplicity, assume a given health intervention leads deterministically to a known, given health outcome. This assumption does not alter the nature of the conclusions reached in this chapter.)

The mechanics of CUA-I can then be summarized as follows:

- In comparing a given health intervention,  $HI_i$ , to the status quo, compute  $D_i = (U(\Delta h_i) \times R_{\text{Threshold}}) - c_i$ . The individual should pursue the health intervention if  $D_i > 0$ .
- In choosing among mutually compatible health interventions,  $HI_1, HI_2, \dots, HI_k$ , all interventions with positive  $D_i$  values should be pursued, i.e., all  $HI_i$  for which  $D_i = (U(\Delta h_i) \times R_{\text{Threshold}}) - c_i > 0$ .

- In choosing among mutually incompatible health interventions, the intervention with the greatest value for  $D_i$  should be chosen.

As in Chapter 3,  $D_i$  can be thought of as the difference between the "imputed" value of the health intervention and its cost. To see this, note that the first of the two terms on the right hand side of the equation that defines  $D_i$  is the imputed value of the health intervention. This is computed as the number of improved health units resulting from the health intervention multiplied by the threshold value associated with each such unit. The second of the terms on the right hand side of the equation that defines  $D_i$  is the cost of the intervention.  $D_i$  is the difference between these two terms, and thus represents the net benefit of the health intervention.

$R_{\text{Threshold}}$  can be viewed as the imputed monetary value the individual places on an expected unit of health. Indeed, as long as changes in health are sufficiently small (see the discussion in Section 10 of Chapter 6),  $R_{\text{Threshold}}$  can be interpreted as the willingness to pay of the individual for an expected unit of health. Then  $D_i$  becomes identical to  $NB_i$ . To see this, note that with this interpretation of  $R_{\text{Threshold}}$ ,  $U(\Delta h_i) \times R_{\text{Threshold}}$  becomes the number of units of health improvement, measured in VNM-utility units, multiplied by the individual's willingness to pay for a VNM-utility unit. In other words,  $U(\Delta h_i) \times R_{\text{Threshold}}$  is the individual's willingness to pay for the health improvement from Health Intervention  $i$ , or  $U(\Delta h_i) \times R_{\text{Threshold}}$  equals  $NB_i$ . Moreover, in the case of a single individual TC<sub>*i*</sub> become simply  $c_i$ . Or,

$$\begin{aligned} NB_i &= TB_i - TC_i \\ &= U(\Delta h_i) \times R_{\text{Threshold}} - c_i \\ &= D_i. \end{aligned}$$

As is easily verified, given this equivalence between  $NB_i$  and  $D_i$ , the mechanics for CUA-I and for CBA become identical.

Thus, for individual decision making, CUA *can* legitimately be seen as an intermediate step toward a CBA. Moreover, there is much to recommend performing an (individual) CBA in the sequential fashion of doing a CUA first. This allows the analyst to assign a dollar value to health as the last step in the analysis. Insofar as assigning a dollar value to health outcomes, i.e., deriving  $R_{\text{Threshold}}$ , may be a particularly controversial or difficult part of the analysis, placing it as the last step has several advantages. It highlights the impact of a given choice of dollar valuation for health on the conclusions of the analysis and prevents this key step from being lost within the intermediate details of the analysis.

### 3.4. The Imperfect Link Between CUA-II and CBA

While CUA-I and CBA are compatible, the same cannot be said necessarily for CUA-II and CBA. In moving from an individual to a societal context, difficulties arise. To see this, consider a set of health interventions which now each affect a group of individuals. As before, suppose there are several health interventions,  $HI_1, HI_2, \dots, HI_K$ . Now, though, let there also be  $N$  individuals,  $1, 2, \dots, N$ , that are affected by the health interventions. Each health intervention,  $HI_i$ , is associated with costs,  $C_i$ , where  $C_i$  represents the total (societal) cost of the health care resources involved. Each health intervention  $i$  is also associated with changes in health of  $\Delta h_{i,n}$ , for each individual  $n$ . Assume for simplicity that the effect of Health Intervention  $i$  on individual  $n$  is deterministic. (The same conclusions are again reached if one allows for a set of possible health outcomes  $\Delta h_{i,n,j}$ , each with probability  $p_{i,n,j}$ .)

To simplify the discussion, assume that these health interventions are mutually exclusive. (The same conclusions about the issues at hand are again reached if the health interventions are mutually compatible.) In applying the principles of CUA-II, the following algorithm should be used:

1. For each of the mutually exclusive health interventions, compute  $D_i = ([\sum_n U(\Delta h_{i,n})] \times R_{\text{Threshold}}) - C_i$ .
2. Choose that health intervention with the largest value of  $D_i$ .

Here,  $R_{\text{Threshold}}$  is the threshold value for CUA-II as discussed in Section 6 of Chapter 7. In applying the principles of CBA, the following algorithm applies:

1. For each of the mutually exclusive health interventions, compute  $NB_i = \sum_n [U(\Delta h_{i,n}) \times R_n] - C_i$ .
2. Choose that health intervention with the largest value of  $NB_i$ .

Here NB stands for net benefit.  $R_n$  is the threshold value for  $R_{\text{Threshold}}$  for individual  $n$ , computed as indicated in Equation (6-5) in Chapter 6.

These two algorithms look the same but upon closer examination they are seen to be different. The problem comes in the first right-hand side term in the equations for  $D_i$  and  $NB_i$ , namely  $([\sum_n U(\Delta h_{i,n})] \times R_{\text{Threshold}})$  and  $\sum_n [U(\Delta h_{i,n}) \times R_n]$ , respectively. It is not generally possible to move from one of these terms to the other. Knowing  $\sum_n U(\Delta h_{i,n})$  and  $R_{\text{Threshold}}$ , the terms on the right hand side of the equation for computing  $D_i$ , does not necessarily allow for the computation of  $\sum_n [U(\Delta h_{i,n}) \times R_n]$ , the term on the right hand side of  $NB_i$ , and vice versa. Thus, knowing  $D_i$  does not necessarily allow for the computation of  $NB_i$  and knowing  $NB_i$  does not necessarily allow for the computation of  $D_i$ .

To see this in the context of a numerical example, suppose there are two health interventions, each of which affects two individuals, A and B. The relevant information for these health interventions is summarized as follows:

- Effects of Health Intervention 1

$$\begin{aligned} U(\Delta h_{1,A}) &= 5 \\ U(\Delta h_{1,B}) &= 10 \\ C_1 &= 20 \end{aligned}$$

- Effects of Health Intervention 2

$$\begin{aligned} U(\Delta h_{2,A}) &= 10 \\ U(\Delta h_{2,B}) &= 1 \\ C_2 &= 10 \end{aligned}$$

- Values for  $R_i$  and  $R_{\text{Threshold}}$

$$\begin{aligned} R_A &= 1 \\ R_B &= 3 \end{aligned}$$

$$R_{\text{Threshold}} = 2$$

- Values for  $D_i$  and  $NB_i$

$$D_1 = 15$$

$$D_2 = 3$$

$$NB_1 = 10$$

$$NB_2 = 12$$

In this example, knowing  $D_1$ , it is impossible to know what  $NB_1$  is without going back to the raw data and computing it. The same holds for  $D_2$  and  $NB_2$ . Similarly, knowing  $NB_i$ , it is not possible to know  $D_i$  without going back to the raw data and computing it.<sup>32</sup>

It is true that much of the same raw information needed to compute  $D_i$ , namely the individual  $\Delta h_{i,n}$  and  $U(\Delta h_{i,n})$  values, are also needed to compute  $NB_i$ , but they are combined in such a fashion in deriving  $D_i$  that it is not possible to compute directly  $NB_i$  from  $D_i$ . Thus, while there will be substantial overlap in the types of information gathered in the two types of analysis, in the end there generally will be no direct link between the conclusions of CBA and of CUA-II. As such, CUA-II--defined as a methodology working with the ratio of costs to aggregate health benefits and with health benefits measured in terms of VNM-utility--cannot generally serve as an intermediate step toward CBA. Indeed, the above numerical example is an extreme instance of this. In that example, CBA recommends Health Intervention 2 over Health Intervention 1 while CUA-II recommends Health Intervention 1 over Health Intervention 2.

### 3.5. Why Do CUA-II and CBA Differ?

Why do CUA-II and CBA differ? It might be thought that the problem lies with the notion of attaching a willingness-to-pay value to health. This, however, is not the difficulty. As discussed above, in principle there is no more difficulty in assigning willingness-to-pay value for a unit of health than there is in calibrating health in terms of VNM-utility.

The heart of the distinction between CUA-II and CBA derives not only from the fact that CBA measures the value of a health impact on a person in terms of that person's willingness-to-pay for health, but also the fact that an individual's willingness to pay for health, as reflected in  $R_i$ , will generally be a function of the wealth of the individual.

---

<sup>32</sup>One could try to link  $NB_i$  and  $D_i$  by taking the value for  $NB_i$  and deriving a value for  $R_{\text{Threshold}}$  such that it would make  $NB_i$  and  $D_i$  the same. This would be done in the following way. Take  $NB_i$ . Add  $C_i$ , yielding  $\sum_n [U(\Delta h_{i,n}) \times R_n]$ . Compute  $[D_i + C_i] / \sum_n [U(\Delta h_{i,n}) \times R_n]$  and define this as  $R_{\text{Threshold}}$ . Having defined  $R_{\text{Threshold}}$  in this fashion, it is possible to use it in the equation for  $D_i$  and yield the "correct" value for  $D_i$ . However, going back and forth from  $D_i$  to  $NB_i$  is completely artificial and does not generalize across health interventions. In general, there is no single value for  $R_{\text{Threshold}}$  that could be computed in this fashion and which would remain fixed across different health interventions. In the numerical example above, this procedure would yield 7/6 for  $R_{\text{Threshold}}$  when using Health Intervention 1 but 13/22 for  $R_{\text{Threshold}}$  when using Health Intervention 2.

Health is generally considered to be what economists call a "normal" good. A normal good is one that an individual purchases more of as his or her income increases. Cars, tickets to the movies, food, and shelter are generally normal goods, as are most other items in a consumer's budget.<sup>33</sup> It is easily shown that normal goods are also goods for which an individual's willingness to pay for a given unit or amount of the good increases the wealthier one is. As applied to health care, this means that the wealthier an individual is, the greater will be his willingness-to-pay for a unit of health.

Equivalently, the wealthier an individual the lower will be his value for  $R_i$ , reflecting the fact that he is willing to reap less health return per dollar than if he were poorer. This in turn means that the outcome of a CBA of a health intervention can well be sensitive to the wealth of the individuals whose health are affected by the intervention. On the other hand, the wealth of an individual does not affect the VNM-utility measure of a given health effect. A CUA of a health intervention will not be affected by the wealth of the individuals whose health are affected by it.

This potential sensitivity of the "value of health" to the distribution of the income of those affected is an important difference between CBA and CUA. There will be no such sensitivity in CUA and there generally will be in CBA. As a result, it may be that CBA will view a health intervention that positively affects the health of wealthy individuals as more desirable in terms of net benefits than a health intervention with identical costs and identical health outcomes but for which those affected are poorer. This is not the case for a CUA.

This link of the potential sensitivity of CBA to the wealth of the individuals affected by a health intervention is illustrated by the numerical example of the previous section. There, CBA recommended Health Intervention 2 while CUA recommended Health Intervention 1 because Health Intervention 2 particularly helps individual A who, because of greater wealth, has a higher willingness-to-pay for health than does individual B. This is enough to have Health Intervention 2 appear more attractive than Health Intervention 1 from a CBA perspective but not from a CUA perspective, where the health impacts to persons A and B are valued the same per unit of health.

This distinction between CUA and CBA reflects a deeper and more fundamental distinction between the two approaches. CBA emerged as a form of analysis with direct links to traditional, neoclassical welfare economics. As such, it is inextricably linked to a basic concern with economic efficiency. A health intervention with positive net benefits is recommended by CBA because of the fact that (with proper monetary transfers, if necessary) it is possible to make all individuals better off than without the intervention. A given health intervention can have positive net benefits, in turn, only if there is a "market failure"—an instance where the private market fails to achieve efficiency on its own. As in neoclassical welfare economics more generally, government intervention in the economy is justified from an efficiency perspective only when a market failure occurs and CBA can be seen as a tool to identify if there is a market failure in a particular area (and to gauge its magnitude).

This perspective may be perfectly fine for some purposes in health care. There are government health programs that can clearly be justified in terms of economic efficiency and market failure. Inoculation programs for children, for instance, can be justified as a response to a nonexcludable public good as can many sanitation programs and basic health research activities.

---

<sup>33</sup>Normal goods can be contrasted with "inferior" goods, which an individual buys less of as his/her income increases. Second-hand clothes, cheap white wine, and pork and beans are examples of goods that most individuals buy less of the wealthier they are.

But, these are not the only types of government health care programs in existence nor, monetarily, are they even the most prominent. More typical of government health programs are programs that are redistributive in nature. Medicare, Medicaid, veterans' health benefits, and women-infant-children programs come quickly to mind as examples. If these programs are indeed viewed as fundamentally redistributive, it may make little sense to base their evaluation on a methodology such as CBA, which is geared strictly to efficiency concerns and, in many ways, purposely ignores distributional impacts.

For these programs, CUA becomes an alternative evaluation approach. CUA ignores differences in the value that individuals place on health that can be attributed to differences in wealth.<sup>34</sup> Instead, it adopts more a constitutional-choice or utilitarian perspective such as put forth in Chapter 7. Individual health status is viewed as comparable across individuals and each person's health is viewed as equally important as every one else's. CUA, as has been seen, becomes a methodology for evaluating how to maximize the aggregate health benefits so defined from health care resources.

While CUA-II thus cannot generally be viewed as an intermediate step to a CBA, it should be noted that this is not to say that one approach is necessarily right and the other wrong. CUA and CBA, even when not directly linked to one another, can still be seen as complementary tools of analysis, each providing different information to a policymaker. If a government health program is motivated by traditional market failure arguments, in which the government's objective is to promote economic efficiency, then the information provided by a CBA will probably be most relevant in evaluating alternative health interventions. On the other hand, if the government objective comes closer to maximizing the aggregate health from a given level of resources—where the wealth and other features of the individual are not deemed of relevance—then CUA-II will be a particularly valuable tool of analysis.

### 3.6. Conditions When CUA-II and CBA Are Compatible

Having argued that there is not generally a direct link between CUA-II and CBA, it is important to emphasize several special situations in which a direct link can be made. Three such situations are identified.

The first occurs when all of the  $R_n$  values are the same. Then, the computation of  $NB_i$  can be written:

$$\begin{aligned} NB_i &= \sum_n [U(\Delta h_{i,n}) \times R_n] - C_i \\ &= \sum_n [U(\Delta h_{i,n})] \times R_n - C_i \\ &= \sum_n [U(\Delta h_{i,n})] \times R_{\text{Threshold}} - C_i, \\ &= D_i, \end{aligned}$$

where  $R_{\text{Threshold}}$  is the (common) value for  $R_n$ . In this situation, it is straightforward to go from  $NB_i$ , the output of a CBA, to  $D_i$ , the output of a CUA-II, and vice versa. Thus, when all of the  $R_n$  values are the same, CUA-II *can* be viewed as an intermediate step in a CBA.

---

<sup>34</sup>Going one step further, CUA can be thought of as an evaluation mechanism for health care resource allocation that is fully independent of the distribution of income in society. To make this argument, it is necessary that the price or opportunity cost of particular health care resources be independent of the distribution of income. As a first approximation, this would seem a reasonable assumption.

How realistic is the assumption that  $R_n$  will be the same across individuals? Insofar as health is a normal good, it may seem that this assumption is not too realistic. However, this conclusion may be overly hasty. There may well be health care interventions for which the individuals involved are of the same or very similar wealth and/or socioeconomic status. In this case, there is no conceptual difficulty in linking CBA and CUA-II.

In the second special case, it may be that a government program affects the health of individuals of different wealths and incomes but there may be no correlation between an individual's value of  $R_n$  and the level of health benefit he receives from a government program. To see that this allows for a direct link between CUA-II and CBA, let  $R$  be the average or expected value of  $R_n$  across individuals, i.e.,  $R = E(R_n)$  where  $E(.)$  is the notation for statistical expectation. Under this assumption of no correlation between  $R_n$  and the level of health benefit received, the computation for  $NB_i$  becomes:

$$\begin{aligned} E(NB_i) &= E(\sum_n [U(\Delta h_{i,n}) \times R_n]) - C_i \\ &= E(\sum_n [U(\Delta h_{i,n})] \times E(R_n)) - C_i \\ &= E(\sum_n [U(\Delta h_{i,n})] \times R) - C_i, \\ &= E(D_i), \end{aligned}$$

where the lack of correlation allows for the derivation of the second line from the first. Thus, in this special case too, it is possible to go from  $D_i$  to  $NB_i$ , and CUA-II can be viewed as an intermediate step to a CBA, at least in a stochastic sense in which  $NB_i$  and  $D_i$  are viewed in expected value terms.

The third special case arises when one chooses to modify CBA so that it is compatible with CUA-II. In particular, one may reject having the outcome of a CBA depend on the wealth of the individuals affected by the health intervention under consideration. An attempt may be made to separate willingness to pay from some notion of ability to pay. This might be done by, say, specifying some prototypical individual—a 35-year-old individual with some given level of wealth—and calibrating for each individual affected by a health intervention how much this prototypical person would be willing to pay for the health effect experienced.

Something akin to this procedure is done in some risk analysis studies in which a value is derived for the monetary value of a statistical life based on the "compensating wage differentials." This compensating wage differential is the additional amount that workers require in wages for additional increments in occupational health risk.<sup>35</sup> These figures are then used to value the health of individuals affected by a health intervention regardless of whether the wealths or individual willingness to pay of these persons is the same as that used in computing the monetary value of a statistical life. Proponents of CUA may have little or no difficulty with this modification. However, to the degree to which it modifies CBA in such a fashion that the links between CBA and welfare economics are loosened, advocates of CBA may feel uncomfortable with this approach.

In summary, while CUA-I and CBA are compatible with one another, the same is not always true when one moves from an individual to a societal context. This inability of CUA-II necessarily to be able to play the role of an intermediate step to a CBA should not be dismissed as a mathematical inconvenience. It reflects a different approach between CUA and CBA to the evaluation of government health programs. At the same time, there are special cases in which these concerns vanish and in which CUA-II can be seen as an intermediate step to a CBA.

---

<sup>35</sup>For an overview of this approach, see, for instance, W. Kip Viscusi, *Risk by Choice: Regulating Health and Safety in the Workplace*, Cambridge, MA: Harvard University Press, 1983.

#### 4. CUA AND DECISION ANALYSIS

Decision analysis (DA) is another evaluative technique widely used in policy settings to analyze individual and societal decisions. DA devotes special attention to the role of uncertainty in decision making. Accordingly, texts in decision analysis typically focus on issues such as risk preferences, methods to reduce risk through diversification, the value of additional information, and so forth. From its earliest development, DA has used VonNeumann-Morgenstern utility theory as its theoretical underpinning. As such, the discussion of utility theory in Chapter 6 is by no means incompatible with DA.

There is another aspect of DA, however, whereby CUA and DA become distinct analytic tools. DA, apart from its special attention to the role of uncertainty in decision making, also allows the decision maker to adopt any goals that the decision maker deems appropriate. In contrast, CUA predefines the goal of the decision maker to maximizing the expected utility of the health/quality-of-life outcomes that result from the allocation of a given amount of resources among alternative health care interventions. It may seem that this goal structure is the natural goal structure for an individual to adopt but, as will be discussed shortly, this is not necessarily so, particularly in the context of CUA-II.

This "relativist" view of the potential goals of the decision maker has been adopted not only in DA but is ascribed by many to "cost-effectiveness analysis" itself. This view of CEA as a means to further any of a number of possible objectives that a decision maker might have is adopted by OTA (1980) in its ten basic steps for conducting CEA/CBA (it considers CEA and CBA as two interlinked forms of analysis). The steps are defined very generally, as in Chapter 1 of this monograph, and wide latitude is given to how each step is carried out. Any study that in some fashion adopts all or most of these steps presumably qualifies as a CEA. This relativist view is echoed by Warner (1983), who writes,

A health care practice is considered to be cost-effective if it is worth the expenditure of the resources required to perform it. The quotation marks around worth emphasize the subjective element in assessing the value of a medical practice and, hence, its cost-effectiveness.<sup>36</sup>

Hatziandreu et al. (1989) reflect this view as well, writing "cost effectiveness is a relative and subjective concept defined as 'having an additional benefit' worth the additional cost."<sup>37</sup>

From this subjective perspective, CEA can be useful for many quite different sets of objectives and goals. Consider a few that have at one time or another claimed to be the type of goals addressed by CEA:

---

<sup>36</sup>Kenneth E. Warner, *Systematic Analysis: Cost-Benefit and Cost-Effectiveness Techniques*, in B. Bloom and S. Berki, eds., *Cost Benefit, Cost Effectiveness, and Other Decision Making Techniques in Health Care Resource Allocation*, 1983.

<sup>37</sup>Evidiki I. Hatziandreu, Jeffrey P. Koplan, Milton C. Weinstein, Carl J. Caspersen, and Kenneth Warner, *A Cost-Effectiveness Analysis of Exercise as a Health Promotion Activity*, *American Journal of Public Health*, 78(11), 1988.

- an individual choosing among health interventions that affect him, concerned only with his own health and out-of-pocket costs,
- a health benefits manager in a firm trying to hold down health care costs to the firm,
- a negotiator for a union trying to provide the best overall health benefits to the employees whom the union represents,
- a Government agency assessing whether mammogram screening of the elderly will reduce the overall level of Medicare payments,
- a Government agency assessing whether mammogram screening of the elderly will provide a sufficiently large return in health benefits to justify the cost,
- a Government agency assessing the benefits of a health intervention to a particularly important constituency (or to a senator who sits on an important appropriations committee),
- a Government agency attempting to promote equal access to medical care,
- a Government agency attempting to provide a "minimal" level of health care to those who cannot afford it themselves,
- a medical equipment manufacturer trying to convince a hospital that its equipment will lower costs to the hospital, and
- a physician choosing which among several procedures will help his patients at the lowest cost.

And the list goes on. From the perspective of a relativist, DA, and CEA as well for those who feel CEA is also relativistic, may serve any and all of these different purposes. This approach is certainly appealing. Why not have CEA serve as many diverse purposes as possible? After all, CEA is ultimately intended to be of use to policymakers and others involved in health care resource allocation and they may have many different legitimate objectives. Why rule out any of these purposes?

This relativist view does, however, lead to some difficulties. For instance, it leads to the term "cost effectiveness" meaning quite different things to different authors. This has led to confusion, ambiguity, and many a footnote in which an author has tried to relate what he or she means by "cost-effectiveness analysis" to what someone else means by that term. In this context, Doubilet, Weinstein, and McNeil (1986) discuss the growing popularization of the term "cost effective" in the medical literature over the past decade. They write,

Examples of its use abound in the medical literature, in drug and equipment advertisements, and in the hallways and lecture halls of medical institutions. Unfortunately, many who assert that a specified medical practice is cost effective fail to couple that statement with appropriate documentation of associated costs and

benefits. Even when these have been computed, the conclusions drawn from them have been based on different interpretations of the term.<sup>38</sup>

This relativist perspective also has led to a lack of cumulativeness in much research involving CEA in health care. It is often very difficult to compare the results of one study with another and sometimes even to reach agreement on the basic interpretation or meaning of a given CEA.

Additionally, lacking a single consensus model, many consumers of CEA distrust the value of given analyses. The more discretion an analyst has in choosing measurement procedures in CEA, the greater the suspicion that judgments have been made in ways that promote the interest of the individual or organization presenting the analysis.

This observation, while not particularly novel, deserves additional discussion. To the degree to which CEA is intended to provide information to guide resource allocation decisions in health care, it inevitably will be used as an advocacy tool among proponents with vested interests in the conclusion of the analysis. As such, CEA's strength lies in its objectivity; i.e., in providing information about the relative value of competing claims for scarce health care resources, CEA's effectiveness rests largely on the degree to which it can be viewed as an objective tool of analysis.

There are, of course, many ways to define "objective." But, a sufficient definition for present purposes is that an analysis is objective if it can be performed by groups or individuals with different vested interests and lead in each case to the same conclusions. If discrepancies arise, one should be able to resolve them based on agreed-upon rules for measurement and interpretation. It is fair to say that, based on this definition, many CEAs in the current literature do not fully achieve this objectivity and a consensus model for conducting CEA can therefore serve a useful function in this regard.

An analogy may be made to the concept of "accepted accounting practices" in financial reporting by firms. The strength of an audited financial report to those who rely upon it to verify the financial condition of a company is that the financial report is constructed according to a set of predetermined rules that allow comparability over time and across firms. An auditing of the financial report of a given company by different auditors should result in the same conclusions. If the same does not hold true for a tool such as CEA, the claim that a given health intervention is "cost effective" carries no reliable meaning at all. In the extreme, a given CEA may be viewed with the same skepticism as a financial report that freely chooses, from among many options, whichever reporting procedure places the company's finances in the best light.

Clearly, the view of CUA as put forth here, as a way of gauging the cost effectiveness of a health intervention, differs from this subjective view. For the reasons cited, we view CUA as more properly defined in terms of the specific rules and procedures for measurement put forth in the earlier chapters and refined in Part III than would be the case from a subjectivist view. Nonetheless, as in our evaluation of CBA, we do not view this subjectivist approach to evaluating health interventions as wrong and CUA as defined here as right. There is a very legitimate methodology for taking the objectives of a given decision maker, regardless of their parochial or idiosyncratic nature, and helping to determine the course of action that best furthers these objectives. This approach is decision

---

<sup>38</sup>Peter Doubilet, Milton C. Weinstein, and Barbara J. McNeil, "Use and Misuse of the Term 'Cost Effective' in Medicine," *New England Journal of Medicine*, 314(4), p. 253.

analysis or, in its more advanced manifestation, multiattribute utility theory. Both are respected and rapidly evolving fields of research in their own right.

From this view, the application of DA and CEA in its relativist guise to the analysis of health care interventions can be very valuable, and there could certainly be instances in which such an approach would be more valuable than CUA. The points to be made here are that CUA, as defined in this monograph, is neither the same as decision analysis more generally conceived nor is it the same as a relativistic interpretation of CEA. In this regard, CUA is admittedly less flexible than decision analysis. As we have seen, it provides information in the form of a cost-effectiveness ratio, in which the numerator is the social costs of the resources associated with a health intervention and the denominator is the aggregate health impact measured in terms of VNM-utility. It specifies the objective of a policymaker: the maximization of aggregate health impact, measured in terms of VNM-utility, for a given level of health care resources (or, conversely, minimization of the cost of providing a given level of aggregate health care).

While CUA is restricted to furthering this objective, nonetheless there are instances in which the basic mechanics of CUA will be useful for a policymaker even if the objective is not identical to that for CUA. In particular, whenever a policymaker's objective can be framed in terms of the optimization of a single-dimensional output measure subject to a constraint on a single-dimensional input measure, the basic mechanics of CUA will be relevant. For instance, a company's objective in evaluating alternative health insurance packages it offers its employees may be to maximize the aggregate health outcome to employees and their families, measured in VNM-utility, subject to some given level of cost to the company. To simplify for the sake of illustration, suppose that any number of mutually compatible components can be included in the insurance package; however, if all were included, the cost would exceed the money available. Then, the company can use the same methods presented for CUA-II to select components to be included.

On the other hand, suppose a Federal Government policymaker wants to propose a health program that will: (1) be consistent with the President's policy on Federal versus State responsibilities for health care; (2) be acceptable to key Congressmen on several important subcommittees; (3) not overlap with the jurisdiction of another Federal agency; (4) improve health; (5) not add to the Federal deficit. These may be the real constraints and valid objectives of the policymaker. In the face of uncertainties about program effects, costs, political reactions, and so forth to various proposals, decision analysis may have much to offer this policymaker. However CUA, in anything like the form presented here, will not.

Or, to take a less extreme example, suppose a Government agency wants to assess the desirability of a potential screening program on the basis of whether it saves the Government health care costs by reducing future illness sufficiently to offset the cost of the screening. This may (or may not) be a valid criterion for the Government agency in assessing the screening program, but it cannot be expressed as a constrained optimization problem and, again, the mechanics and approach of CUA will not be useful.

None of this is meant to indicate that policymakers in health care must care only about the objectives embodied in CUA. But, it is to suggest that CUA is not synonymous with decision analysis. At the same time, as long as the objectives of a policymaker are to maximize some single-dimensional measure of the output of a health intervention subject to a single-dimensional measure of the inputs involved, then the use of a cost-effectiveness ratio of inputs to outputs for each health intervention will be of value in analyzing their desirability. Moreover, the basic mechanics of CUA as presented

in Part I will be applicable. If the objective of the policymaker cannot be represented in this fashion, however, the tools of CUA will be of correspondingly less value.



## **PART III**

### **ISSUES IN MAKING CUA OPERATIONAL**

Part II of this monograph provides a basic framework for CUA. In one sense, this, in conjunction with the definition of CUA and the mechanics presented in Part I, accomplishes the monograph's mission. However, in actually performing a CUA there are many practical complications that can arise. Part III addresses some of these.



## CHAPTER 9

### COMPLICATIONS WHEN THERE IS MORE TO LIFE THAN HEALTH

#### 1. INTRODUCTION

The analysis in Part II assumed an individual's overall utility to be a function only of his health,  $H$ , and material consumption,  $M$ .<sup>40</sup> For now, the economists' term "overall utility" can be considered synonymous with "health-related quality of life," a term popular in the cost-effectiveness literature.<sup>41</sup> In reality of course, there is more to overall utility, or quality of life, than health alone or even than health and money. This fact helps make life worth living even in poor health, but it does complicate CUA in health care. Overall utility must be considered now as a function of multiple attributes, only two of which are health and material well-being.

#### 2. THERE IS MORE TO LIFE THAN HEALTH

##### 2.1. What Are The Attributes of "Overall Utility"?

There is no definitive list of the attributes of overall utility, but several deserve special mention. Neoclassical economics, particularly in its analysis of labor supply, posits that an individual gets utility from material well-being as reflected by consumption of market goods,  $M$ , and from leisure,  $L$ . Leisure, as defined here, includes more than simply relaxation but more generally any time that one has for one's own use, even if that use is washing the dishes rather than lying on the beach.

Some analyses would go further and divide leisure itself into two components. An example is provided by several recent<sup>42</sup> cost-of-illness analyses (COIAs). In assessing the economic impact of an illness, many recent COIAs consider not only the lost earnings that result from an illness, but also the losses incurred by homemakers in labor devoted to the household. Homemaker losses are valued in different ways, including for instance estimating the cost of purchasing each of the household labor services that are lost due to the illness. Such studies implicitly distinguish between types of leisure. One type of leisure is leisure that directly adds to personal utility, e.g., sleeping, watching television, and playing golf. These are the kinds of activities usually associated with "leisure" and can be termed "self-consumption leisure." Self-consumption leisure can be distinguished from activities which in

---

<sup>40</sup>In prior chapters, material consumption was denoted by  $C$ . It is convenient here to represent material consumption as  $M$  and to distinguish it from "total costs." Previously, when it was assumed that the only inputs to a health intervention were health care resources this distinction did not matter insofar as the only costs to an individual from a health intervention were exactly equal to the lost material consumption. Now, however, some of these "costs" can come in the form of things like leisure time, which may not represent direct decrements to material consumption.

<sup>41</sup>Here, the term "overall" in "overall utility" indicates the VNM utility of an individual as a function of an entire description of the state of the world experienced by the individual. This would include an individual's health but also everything else that might affect his utility. This is in contrast to a more narrow measure of utility such as that part of overall utility that is due to an individual's health state alone.

<sup>42</sup>The "first generation" COIA's took account only of the impact of an illness on market labor force participation. Only beginning in the 1980s did many COIA's begin to measure the impact of an illness on non-market labor activities as well.

some sense are work within the home, such as cleaning, certain child care tasks, cooking, and so forth. Such time is spent substituting work in the home for what otherwise would be purchased. This type of leisure can be termed "uncompensated work in the home," where the term "uncompensated" means that, unlike work performed in a market setting, no monetary wage payment is received.

This orientation is compatible with recent literature on household production functions,<sup>43</sup> which also argues that some activities are performed in the home not for their own innate enjoyment but to produce other goods. These other goods could be purchased outside the household if not produced by the household itself. From the point of view of these literatures, then, leisure can be divided into two types,  $L_s$  and  $L_u$ .  $L_s$  denotes self-consumption leisure and  $L_u$  denotes leisure that is spent performing uncompensated work in the home.

A somewhat different perspective on overall utility is provided by the health psychology literature. Over the past 15 years, substantial effort has been made to measure the overall quality of life of an individual. Examples include the Sickness Impact Profile, the McMaster Health Index Questionnaire, the Nottingham Health Profile, the Psychological General Well-being Index, the General Health Rating Index, and the Quality of Well-being Index. While these scales differ in some details, they do provide a consensus about the generic components of the quality of life, including three factors in particular. One is health. The second is social interactions, including family life, friends, social support networks, and so forth. The third is material well-being.

Two of these, health and material well-being, overlap with the attributes from neoclassical utility theory. With social interactions,  $S$ , as a new variable, we get the following specification for the attributes in quality of life:

$$(9-1) \quad U = U(H, L_s, L_u, S, M).$$

This functional notation simply means that, in some fashion or other, overall utility or quality of life is a function of health, self-leisure, leisure spent performing uncompensated work in the home, social interaction, and consumption.

## 2.2. Direct and Indirect Costs

Health care interventions involve costs. In the CUA-I analysis of Chapter 6, these costs were assumed to consist only of health care resources. Every dollar spent on health care resources was, in turn, one dollar less to spend on other material consumption. In reality, the relationship between a health intervention and material consumption can be more complicated. For instance, one's health can influence how much one is able to work and/or one's productivity. This can affect one's income from labor and, thus, will also affect material consumption,  $M$ .

Taking this complication into account, it is typical in COIAs and other health care analyses to define two types of costs, direct costs,  $C_D$ , and indirect costs  $C_I$ . Direct costs include medical, clinical, and other health care resources that are inputs to the health intervention, e.g., drugs, cost for services of health professionals, laboratory costs, and so forth. Indirect costs are the lost wages and income in

---

<sup>43</sup>See, e.g., Mark R. Rosenzweig and Paul T. Schultz, Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight, *Journal of Political Economy*, 91 (5), October 1983.

the future that may be experienced from an illness due to lower productivity, greater absenteeism, and so forth.

### 2.3. CUA in A Multiattribute Context

In this multiattribute context, a given health intervention may lead to a change in all of the components in Equation (9-1). For instance, health might affect one's social functioning, one's ability to enjoy leisure, or one's ability to work and make money. A health intervention also may involve more than monetary costs. For instance, an exercise program involves not only some degree of monetary costs (to purchase proper equipment, etc.) but also requires the individual to devote time that could be "consumed" doing other activities.

In this multiattribute situation, how should an individual determine if a given health intervention should be adopted? Let the subscript "wo" denote the value of these attributes without the health intervention and let the subscript "w" denote the value of the attributes with the health intervention. Then the individual will judge the health intervention to be desirable compared to the status quo if (and only if):<sup>44</sup>

$$(9-2) U_w > U_{wo},$$

where

$$(9-3) U_w = U(H_w, L_{S,w}, L_{U,w}, S_w, M_w),$$

$$(9-4) U_{wo} = U(H_{wo}, L_{S,wo}, L_{U,wo}, S_{wo}, M_{wo}),$$

and

$$(9-5) M_w = M_{wo} - C_D + C_I.$$

Equation (9-5) indicates that material consumption after a health intervention is determined by first subtracting the direct costs of the health intervention from what  $M$  would be without the health intervention and then adding indirect costs. Consistent with this equation, material consumption could be greater with a health intervention than without. For example, more could be saved in averted lost wages from the health intervention than the direct costs of the health intervention.<sup>45</sup>

Does CUA-I remain well defined for the decision problem represented by Equations (9-2) through (9-5)? The answer is yes. Without going through all the mathematical derivations here, it can be shown that, following the principles of VNM-utility theory, an individual will continue to evaluate a given health intervention relative to the status quo in terms of a cost-effectiveness ratio.

---

<sup>44</sup>This notation assumes that the health intervention has a known, deterministic effect on the attributes in  $U(.)$ . If this is not true, the following equations can be modified straightforwardly, with the individual's objective now being to maximize expected utility.

<sup>45</sup>Given that Equation System (9-2) to (9-5) is from the individual's perspective, Equation (9-5) also assumes that the individual is responsible for direct costs and bears the indirect costs. This may not always or even generally be true. It may be that health insurance picks up some or much of the direct costs and that the individual's employer bears the indirect costs. If such cases,  $C_D$  and  $C_I$  would have to be modified to reflect only those costs borne by the individual.

### 3. A FUNDAMENTAL EQUIVALENCE RESULT FOR CUA-I

The conclusion from Section 2 that CUA remains applicable for evaluating health care interventions in a multiattribute context is important and practical. It may seem to resolve all issues that could arise in extending CUA to a multiattribute setting. However, an immediate complication arises. A given effect of a health intervention (whether an input or an output) might be captured either in the numerator of the cost-effectiveness ratio as a cost or in the denominator as an impact on the quality of life. Which should it be?

For instance, take a simple case in which there are no causal linkages between any of the components of overall utility (i.e., changes in health do not affect social interactions, the amount an individual can work, etc.) and in which all health care resource costs are paid by the individual. In this situation, the only costs of the health intervention are direct costs,  $C_D$ . For simplicity assume that overall utility, as specified in Equation (9-1), is additively separable not only in consumption  $M$ , as assumed throughout, but also leisure. Also for simplicity, let leisure not be differentiated into self-consumption and home production and let the role of social interactions,  $S$ , be ignored. Then, overall utility can be written:

$$(9-6) U = U_1(H) + U_2(M) + U_3(L).$$

Let  $\Delta H$  be the change in health from some potential health intervention and let  $U_1(\Delta H)$  be the change in overall utility from that level of change in health. To this simple case, add only one complication: the health intervention requires the individual to devote time to it and this time otherwise would have been available as leisure.

If the health intervention consumed only money and affected only health, it would be evaluated from a CUA perspective by computing the ratio  $C_D/U_1(\Delta H)$ . The health intervention would be cost effective if  $C_D/U_1(\Delta H)$  were less than  $R_{\text{Threshold}}$ . Now, however, we also must account for the impact of the health intervention on leisure. One way to do this is to include its impact in the denominator of the cost-effectiveness ratio. This denominator would now capture not only the (positive) impact of the improvement in health from the health intervention but also the (negative) impact of the time devoted to the health intervention. Both would be measured in terms of their impact on the individual's overall utility or quality of life.

Letting  $\Delta L$  be the leisure consumed by the intervention and  $U_3(\Delta L)$  be the change in overall utility caused by this change in leisure (where  $U_3(\Delta L)$  will be negative), then this cost-effectiveness ratio would be:

$$(9-7) C_D/[U_1(\Delta H) + U_3(\Delta L)].$$

Let  $R_{\text{Threshold}}$  be defined as in Chapter 6 as  $1/Z$ , where  $Z$  represents the increase in utility that results from the availability of an incremental dollar, and is defined mathematically as it was in Equation (6-5) of Chapter 6:

$$(9-8) Z = [\partial U / \partial M] (\equiv 1/R_{\text{Threshold}}).$$

The cost-effectiveness ratio in Equation (9-7) can be used to assess the desirability of a health intervention in the normal fashion. The logic of the CUA-I analysis of Chapter 6 remains intact. That is, a health intervention can be judged from a CUA-I perspective in terms of whether its cost-

effectiveness ratio, defined in Equation (9-7), is less than or greater than  $R_{\text{Threshold}}$ . So, the health intervention will be cost effective if:

$$(9-9) \quad C_D / [\Delta U_1(H) + \Delta U_3(L)] < R_{\text{Threshold}} = 1/Z,$$

or if:

$$(9-10) \quad ZC_D < \Delta U_1(H) + \Delta U_3(L).$$

But suppose that one approaches the analysis differently and chooses to modify the numerator of the cost-effectiveness ratio instead of the denominator. Suppose that rather than measuring leisure in terms of quality of life, a direct monetary value is attached to leisure. How might this be done and how would the results of this alternative procedure affects the conclusions of the analysis?

To begin, what is meant by the monetary value of leisure? In this context, the monetary value to be attached to leisure is the individual's willingness to pay for the leisure. It is the most the individual is willing to pass up in monetary payment in order to have leisure for personal use rather than have someone else direct it. To illustrate this concept, suppose an individual, Joe, is offered money by his neighbor to mow his neighbor's yard. This task takes one hour. He would rather use this hour for his own use to watch TV. His neighbor offers him \$10 to mow the lawn and Joe defers. If he were offered \$1,000, however, Joe would be happy to accept. In fact, his cutoff point is, say, \$14.38. For any price above \$14.38 Joe will agree to mow his neighbors yard while for any price less he chooses not to. In this case, \$14.38 would represent the monetary value that Joe attaches to his marginal hour of leisure or, equivalently, his willingness to pay for the marginal hour of leisure.<sup>46</sup>

In practice, leisure is often valued monetarily in terms of an individual's wage rate. The wage rate is often used as a monetary measure of an individual's leisure because basic microeconomic theory suggests that, at least under simple stylized conditions, an individual has a willingness to pay for the marginal unit of leisure equal to his wage rate.

Let  $w$  denote the monetary value of a unit of leisure. Then this alternative approach to CUA-I would use as the numerator in a cost-effectiveness ratio not direct costs alone,  $C_D$ , but  $C_D + (\Delta L \times w)$ , the cost of the health care resources involved in the health intervention plus the monetary equivalent of the leisure used up in the health intervention. The denominator of the CUA ratio is not the overall change in quality of life from the intervention, including the impact on  $H$  and  $L$ , but only  $U_2(\Delta H)$ , the impact of the intervention on health. The resulting cost-effectiveness ratio would be:

$$(9-11) \quad [C_D + (\Delta L \times w)] / U_2(\Delta H).$$

---

<sup>46</sup>It may seem odd to describe this figure as Joe's "willingness to pay" since he is not literally paying anything. However, consider Joe going to the store and buying an apple for \$1.00. Joe clearly is willing to pay \$1 (or more) for the apple. After buying the apple, he has \$1 less than he otherwise would but has an apple to consume. By the same logic, by passing up \$14 to work on his neighbor's yard, he has \$14 less in his pocket that he otherwise would but has an hour of time to consume for himself. In this sense, the term "willingness to pay" is correctly used when referring to leisure even though no out of pocket payment is made.

What relationship will there be, if any, between the conclusion to a CUA based on cost-effectiveness ratio (9-11),  $[C_D + (\Delta L \times w)]/U_2(\Delta H)$ , and the conclusion based on cost-effectiveness ratio (9-7),  $C_D/[\Delta U_1(H) + \Delta U_3(L)]$ ?

Without going into details here, it can be shown mathematically that the two ratios lead to identical conclusions. Thus, when a health intervention involves leisure as well as direct resource costs one can construct a cost-effectiveness ratio in either of two ways: (1) with costs in the numerator, health and leisure in the denominator, and leisure measured in terms of quality of life; or (2) costs and leisure in the numerator, health in the denominator, and leisure measured in monetary terms. The same conclusion about the desirability of the health intervention will be reached in either case.

This result does not rely upon the attribute being leisure. Impacts on other attributes can be handled in either of the two ways, and the conclusions about the cost effectiveness of a health intervention will be the same.

Note that this result does not indicate that the cost-effectiveness ratio will be the same regardless of which approach to computing the ratio is used. The two approaches give different numerical quantities. But the conclusion will be the same; i.e., the cost-effectiveness ratio will be on the "same side" of  $R_{\text{Threshold}}$  regardless of which computation approach is used. Moreover, it can be shown that all of the mechanics for conducting CUA as presented in Part I continue to be valid and to give the same conclusions regardless of which computation approach is used.

Thus, when attributes of quality of life other than health and costs are involved, directly or indirectly, in the assessment of a health intervention, CUA-I provides latitude as to whether these attributes are "costed out" and included in the numerator of a cost-effectiveness ratio, or included directly in terms of their impact on the quality of life of the individual involved.

#### 4. SOME CAVEATS TO THE FUNDAMENTAL EQUIVALENCE RESULT

There are several qualifications to the above. First, this analysis assumes that the magnitude of the costs and other components of utility that are encompassed in the numerator are relatively small compared to an individual's overall wealth. If this is not the case, then the analysis encounters the type of "lumpiness" difficulties mentioned previously.

For example, suppose that  $Z (= \partial U / \partial M)$  is .3, indicating that the individual receives .3 units of utility from the marginal dollar devoted to consumption goods. This gives an  $R_{\text{Threshold}}$  value of  $1/.3$ , or 3.33. Suppose a health intervention is available that has very large costs and very large health effects and that its cost-effectiveness ratio is 3.0. Under the procedures as outlined above, this health intervention is judged cost effective. However, the costs associated with the health intervention may be so large, that the amount of (nonhealth) consumption the individual is able to enjoy may be significantly compromised. This will make the value of  $Z_C$  increase. Suppose  $Z_C$  increases to .4. This value of  $Z$  implies a value of  $R_{\text{Threshold}}$  of 2.5 after the health intervention is adopted. Thus, after it has been adopted, the health intervention, with its cost-effectiveness ratio of 3, appears not to be cost effective.

This type of difficulty does not arise when the costs associated with the health intervention are sufficiently small that  $Z_C$  does not change appreciably before and after a health intervention is adopted. Even when this lumpiness occurs, there are ways to handle it easily without having to abandon the basic principles of CUA. However, the situation does become somewhat more complex and is not explored here in further detail.

A second qualification to the above derivation of the fundamental equivalence result concerns the assumption that an individual sets  $Z_i$  equal across all utility categories. This assumes not only a good deal of rationality on the part of the individual but also that the individual has substantial flexibility in the allocation of resources. It may well be, however, that an individual is not entirely free to allocate resources according to personal choice.

Labor and leisure are good examples of this. In the framework set forth above, an individual chooses to devote time to market-work up to the point at which he is just indifferent between supplying another hour for work and receiving the wages from that hour of work, and devoting that hour instead for his own use as leisure (either in the form of self-consumption leisure or uncompensated work in the home). In this situation, the incremental utility the individual receives from the wages from the last hour worked and the incremental utility from the last hour spent as leisure are equal.

However, in reality, labor markets are not always the smooth, flexible mechanisms suggested in elementary economics texts. For example, individuals may have little flexibility in how many hours they work, which days of the week they work, and so forth. In these situations, the incremental utility the individual receives from the wages from the last hour worked and the incremental utility from the last hour spent as leisure may not be equal to one another.

For instance, if an individual could choose with complete flexibility how many hours per week to work, he might choose to work 50 hours per week if the wage rate was \$20 per hour. However, the individual may in fact face a "take it or leave it" option to work 40 hours per week at \$20 per hour or 0 hours per week, and chooses the former. In this case, he would rather work more if he could. That is, he gets less incremental utility from the marginal unit of leisure than he would get if he used this marginal unit of leisure for market work instead.

Another example of this type involving labor markets is the wedge that taxes drive between the wage rate and the amount of income a person receives. A person may make \$20 per hour but only take home \$15 per hour after taxes are taken from his paycheck. His decision about how much to work is a function purely of the amount of money he actually takes home, namely \$15 per hour, not his nominal wage rate of \$20 per hour.

As a result of these factors, there may be a divergence between the utility of another hour devoted to leisure and another hour devoted to work. As Hatziaandreu et al. (1988) write, "both theory and empirical analysis suggest that the opportunity cost of the marginal leisure hour is less than the wage rate."<sup>47</sup> That is, the incremental utility that would be gained if the individual received another hour's worth of wages (without taxes being taken out) is greater than the incremental utility from using that hour as leisure instead.

Fortunately, the underlying logic of the analysis remains intact. Again omitting the mathematical details here, it can be shown that evaluation of a health intervention remains the same. Now, however, a unit of leisure must be valued monetarily at a "shadow price" reflecting its true opportunity cost to the individual.

---

<sup>47</sup>Hatziaandreu, op. cit., p. 1420. Put differently, this means that an individual would agree to work another hour at his present pay rate (with no taxes taken out) if he could.

## **5. COMPLICATIONS MOVING FROM CUA-I TO CUA-II**

The fundamental equivalence result derived in Section 3 pertained to an individual. In moving from the individual perspective of CUA-I to the societal perspective of CUA-II, complications arise. Unlike the conclusion for CUA-I, there is no guarantee that the results of a CUA-II will be insensitive to the choice of which components to measure in the numerator of a cost-effectiveness ratio and which to measure in the denominator.

The difficulty encountered is similar to that faced in viewing CUA-II as an intermediate step to CBA, which, as seen in Chapter 8, is true only in special circumstances. The problem stems from the fact that economic costs—captured in the numerator of a cost-effectiveness ratio—are measured in terms of opportunity costs and, ultimately, in terms of the willingness to pay of those affected. People, though, can have different willingness to pay for something even if it provides them with the same amount of utility. To see the implications of this, consider the following example.

Suppose that Health Intervention A improves health, requires some large amount of time on the part of the individuals involved, and requires little in direct cost. Assume the time involved would have been consumed as leisure. Now consider an alternative health intervention, Health Intervention B, that has the same effect on health, requires very little time on the part of the individuals involved, and requires much more in direct cost. There may be modest differences among individuals of different wealths in the utility they attach to incremental units of leisure but substantial differences in their willingness to pay for incremental units of leisure.

If leisure is "costed out," valued for instance in terms of the wage rate, and put in the numerator of a cost-effectiveness ratio, individuals who are wealthy and have a high willingness to pay for leisure will have a large impact on the cost-effectiveness ratio. As a result, Health Intervention B, which requires little expenditure of leisure, will appear attractive. On the other hand, if leisure is not costed out but measured in terms of its impact on quality of life and, therefore, in utility units, the impact of the wealthier individuals on the resulting cost-effectiveness ratio will be much less pronounced. As a result, Health Intervention A, which economizes on direct costs at the expense of more leisure being required, will appear attractive. It is possible that measuring leisure as an economic cost in the numerator of a cost-effectiveness ratio will lead to Health Intervention B being preferred to Health Intervention A while measuring leisure in terms of quality-of-life impact in the denominator of the cost-effectiveness ratio will lead to Health Intervention A being preferred. While this example is stylized, it points out that in CUA-II the results of the analysis can be affected by which inputs and outputs of a health intervention are captured in the numerator of a cost-effectiveness ratio and which in the denominator.

## **6. WHAT SHOULD BE IN THE DENOMINATOR OF A COST-EFFECTIVENESS RATIO?**

### **6.1. Introduction**

In light of the above analysis, how should a cost-effectiveness ratio be defined? Which effects should be "costed out" in the numerator and which measured in terms of impact on utility in the denominator? There is little agreement in the literature on this question. At one extreme, only health, narrowly defined, is measured in the denominator. All other inputs or outputs are costed out and included in the numerator. Advocates of this perspective might feel comfortable defining the units of a cost-effectiveness ratio as "cost per wellness years (\$/WY)," where the term "wellness" suggests an outcome measure in the denominator of the cost-utility measure which considers health in a relatively circumscribed sense (Patrick, Bush and Chen, 1973; Kaplan and Anderson, 1988).

At the other extreme, some analyses place all aspects of a health effect, except the direct costs of the health intervention, in the denominator of a cost-effectiveness ratio. Only direct health care costs are included in the numerator. Advocates of this perspective might feel comfortable defining the units of a cost-effectiveness ratio as "cost per quality adjusted life years (\$/QALYs)," a term introduced by Weinstein and Stason (1976) and by Bush, Fanshel, and Chen (1983). Here, all the noncost impacts of a health intervention on an individual's life are captured in the denominator of a cost-effectiveness ratio, leaving only some measure of direct costs in the numerator.

While both of these approaches to measuring a cost-effectiveness ratio appear in the literature, by and large QALY appears to be the more common approach, having been adopted by the U.S. Congress' Office of Technology Assessment (1979).<sup>48</sup> Kaplan and Bush (1982) argue, however, that QALY has a connotation which may be construed as overly broad and extending beyond the realm of health. Instead, they recommend wellness years for use in policy analysis of health care programs. Moreover, even among studies using QALY, the precise definition of quality of life is often left vague. Campbell, Converse, and Rogers (1976) describe quality of life as a "vague and ethereal entity, something that many people talk about but which nobody very clearly knows what to do about."<sup>49</sup> Deniston et al. (1989), reviewing the literature since Campbell et al., write that "subsequent literature suggests that we are far from agreement on how to measure the concept."<sup>50</sup>

For simple health interventions involving only direct costs and affecting only health, narrowly conceived, the differences between using \$/WY and \$/QALY in a cost-effectiveness ratio disappear. However, for more complex health interventions, the issue of which effects to capture in the numerator of a cost-effectiveness ratio and which in the denominator involves fundamental measurement questions for CUA. The literature of CUA and CEA seldom addresses this matter directly. As a result, the overlapping issues of how to define health, how to define quality of life, and what to include in the numerator and denominator of a cost-effectiveness ratio are probably the most confused in CUA of health care today. Indeed, it is probably not unfair to characterize the literature as adopting more or less every possible mathematical permutation for which effects to capture in the numerator and which in the denominator. Moreover, there often is a lack of consistency in defining the terms in the numerator and denominator of the cost-effectiveness ratio, leading some effects to be omitted entirely and others to be counted twice.

## 6.2. Six Measures of Quality of Life

From the perspective of CUA-I, the notion of dollars per quality-adjusted life-year may at first glance seem inherently ambiguous. If one defines quality of life (QOL) broadly to consider all impacts on an individual, and if this measure of quality of life is captured in the denominator of a cost-

---

<sup>48</sup>See, e.g., Office of Technology Assessment, *A Review of Selected Federal Vaccine and Immunization Policies: Bases on Case Studies of Pneumococcal Vaccine*. Washington, DC: U.S. Government Printing Office, 1979.

<sup>49</sup>A. Campbell, P.E. Converse, and W. L. Rodgers, *The Quality of American Life*, New York: Russell Sage Foundation, 1976, p. 471. Cited in O. Lynn Deniston, Patricia Carpentier-Alting, Jill Kneisley, Victor M. Hawthorne, and Friedrich K. Port, Assessment of Quality of Life in End-Stage Renal Disease, *Journal of Health Services Research*, 24(4):555-78, October 1989.

<sup>50</sup>Lynn Deniston, Patricia Carpentier-Alting, Jill Kneisley, Victor M. Hawthorne, and Friedrich K. Port, Assessment of Quality of Life in End-Stage Renal Disease, *Journal of Health Services Research*, 24(4):555-78, October 1989, p. 556.

effectiveness ratio, then there is literally nothing left for the numerator. Even defining the numerator in the most narrow terms as direct costs which the individual experiences, it is not at all clear why these, too, do not affect his quality of life by diminishing material consumption and, thus, belong in the denominator.

To resolve this semantic issue, the term "quality of life" can be viewed in this section as encompassing all those things that are captured in the denominator of a cost-effectiveness ratio and to exclude those factors that are counted as costs and captured in the numerator of a cost-effectiveness ratio. Using the multiattribute framework presented in this chapter, this leads to six definitions of quality of life that can be posited, each of which could be used in determining the components of the numerator and denominator of a cost-effectiveness ratio. These are listed as follows:

- QOL Definition 1:* Quality of life encompasses only health,  $H$ .
- QOL Definition 2:* Quality of life encompasses health,  $H$ , and social interactions,  $S$ .
- QOL Definition 3:* Quality of life encompasses health,  $H$ , social interactions,  $S$ , and leisure devoted to own consumption,  $L_o$ .
- QOL Definition 4:* Quality of life encompasses health,  $H$ , social interactions,  $S$ , leisure devoted to own consumption,  $L_o$ , and leisure devoted to home production,  $L_p$ .
- QOL Definition 5:* Quality of life encompasses health,  $H$ , social interactions,  $S$ , leisure,  $L$ , and indirect costs,  $C_i$ .
- QOL Definition 6:* Quality of life encompasses health,  $H$ , social interactions,  $S$ , leisure except for the individual's direct time involvement in the health intervention,  $L_D$ , and indirect costs,  $C_i$ .

Note that QOL Definition 6 subdivides leisure into a third category besides  $L_o$  and  $L_p$ . This third category of leisure is the time the individual spends on the health intervention itself. This category of leisure is denoted by  $L_D$ , which stand for the individual's direct time involvement in the health intervention.

In moving from QOL Definition 1 to Definition 5, an increasing number of attributes is included in the denominator of a cost-effectiveness ratio and increasingly fewer in the numerator. Definition 1 is the most narrow of the definitions of quality of life and corresponds most directly to the denominator in a cost-effectiveness ratio measuring only "wellness" or wellness years.

Most measures of health status, however, do not measure health as narrowly as this. Often, health is considered to be inextricably bound with an individual's ability to enjoy social interactions and even to get the most out of leisure activity. For instance, Hall et al. (1989), in a study of elderly patients, indicate that health is perceived as an umbrella concept embodying functional health, cognitive functioning, emotional health, physiological health, social activity within the family, and social activity outside the family. In our terminology, this is to say that health and social interaction (and, possibly, leisure as well) are not additively separable in an individual's overall utility function and that it is not sensible to consider them in isolation. As such, Definition 2 or 3 are, in practice, closer to how health is conceptualized in many measures of health status than is true for Definition 1.

Definition 3 is also generally consistent with the recent generation of COIAs which recognize forfeited home-production activities as a cost of an illness. These analyses implicitly divide leisure into home production and own consumption, considering home production as an indirect cost. Applied to CUA, quality of life becomes everything else in overall utility, as in Definition 3.

Definition 4 is consistent with more narrow (and more traditional) COIAs. Costs from such a perspective include only direct and indirect costs and indirect costs include only lost wages, not lost home production effort.

Definition 5 goes one final step further and only measures direct costs in the numerator of a cost-effectiveness ratio. Impacts of a health intervention on future work ability, and the lost income from it, are captured in the implications of the lost work and lost income on the quality of life of the individual.

Definition 6 is in the spirit of Definition 5. It, however, singles out patient time inputs of a health intervention, defining such time inputs as costs in the numerator of the cost-effectiveness ratio. This would include time waiting at a doctor's office, transportation time to and from a doctor's office, time in the intervention itself (e.g., time on a kidney dialysis machine, time devoted to exercise, time in therapy, etc.).

### **6.3. The Preferred Measure of Quality of Life, and Therefore, of Costs in CUA**

#### **6.3.1. Introduction**

How can one judge which of the above six possible definitions of QOL is appropriate for CUA? From the perspective of CUA-I, any of them might be used. The CUA analysis will reach the same conclusions whichever approach is used as long as there is consistency between the way costs are measured in CUA and the way that quality of life is measured (a consistency requirement that is often honored in the breach in the literature).

Still, even in CUA-I there is an advantage for choosing one measure of QOL and having it be the accepted procedure for defining a cost-effectiveness ratio. This allows for comparability among studies and for a "level playing field" where alternative analyses can be judged against a common standard. Additionally, while any of the above six definitions is theoretically acceptable in CUA, this does not mean that each is equally easy to implement correctly. It thus becomes important to consider which approach is generally going to be the easiest to implement correctly.

In moving from CUA-I to CUA-II, the issue becomes a bit more complex. For practical reasons, a definition for CUA-II should be chosen that is a direct extension of the preferred definition for CUA-I. It is hard to imagine adopting one definition of quality of life for individual analyses and another when applying the analysis to a societal context. This would lead to almost as much confusion among practitioners and consumers of CUA as the current situation in which little attention is paid to the issue at all.

On the other hand, one must be careful in extending the desirable definition of a cost-utility ratio from a CUA-I perspective to CUA-II. As seen, different definitions of the cost-utility ratio in CUA-II can lead to different evaluations and conclusions of the analysis. In essence, items placed in the numerator of the cost-utility ratio are ones for which differences in willingness to pay across individuals due to differences in wealth are judged legitimate influences on the outcomes of the analysis.

In the simple world considered at the beginning of this chapter, in which the only costs are direct health care resource costs and the only impacts are on the health of the individual individual involved, all six definitions become equivalent. If the world were only so simple, so would CUA be in health care. In general, however, health interventions involve more than direct costs, health interventions may directly affect more than health, and health itself may well indirectly effect other components of overall utility.

In these more complex but realistic settings, the definition of quality of life which is adopted affects fundamentally the measurement of the numerator and denominator in cost-utility analysis. However the numerator and denominator are measured, it must be in a mutually consistent fashion. Given, then, that the issue cannot be avoided, which definition of the above six should generally be used in CUA-I? While it may be arbitrary theoretically which definition of quality of life is adopted, there are important practical reasons why one definition might be preferred to another. Four practical reasons are advanced in favor of Definition 6 for defining quality of life in CUA.

### 6.3.2. *Indirect Costs*

The first argument in favor of QOL Definition 6 involves indirect costs. (It pertains equally to QOL Definition 5 as well). Imagine two individuals, Joe and Pete. Joe is retired, wealthy, and plays a lot of golf, around which a good deal of his social interactions revolve as well. Pete is identical to Joe in every respect except he is not retired and not wealthy and plays no golf (although he would like to if he had the time and money). Each experiences a back injury which leaves Pete unable to work and Joe unable to play golf.

The impact of the injury on Pete's future income stream may be relatively easy to gauge. If indirect costs are incorporated in the numerator of a CUA ratio, as in QOL Definitions 1 through 4, typically only this effect would be explicitly accounted for in a CUA. But, this would be wrong for two reasons.

First, Pete, while unable to work, suddenly has an increase in his amount of leisure. If used as self-consumption leisure, this increased leisure may not be as enjoyable to Pete if his back were not injured but it is not necessarily of trivial utility compared to, say, being dead. If not used as self-consumption leisure, it might be used as home-production leisure. Perhaps Pete takes over added household and childrearing duties so that his wife, previously a homemaker, now is able to work. In either case, the value of the additional leisure to Pete must be taken into account and deducted from the monetary implications of lost future income.

Joe may get much less sympathy than Pete. He is no longer able to play golf but this hardly seems of nearly the impact on Pete of lost future wages. But, is this such an easy conclusion? Joe could have worked like Pete but, given his wealth, chose not to. By revealed preference he enjoyed spending his time golfing more than the income he could have received from working. As such, the inability to play golf may well be more of a disability than the inability to earn income. As such, the impact on Pete as captured in a CUA in health care should, if anything, be greater than the impact on Joe.

We know of no CUA that has successfully taken these complications into account. Indeed, one would be very hard pressed to think of one that has even addressed the issue in a very systematic way. Such factors cannot be ignored with impunity, though. The outcome from so doing is erroneous results in which some costs are overstated, e.g., for failing to deduct Joe's offsetting increases in utility from increased leisure, while others—the opportunity cost of Pete's inability to play golf—are omitted entirely. Not only are the results erroneous, but they typically will be biased in a particular fashion.

Specifically, individuals who are not working will tend to be systematically discounted. The impact of health on homemakers, the unemployed, the employed, the elderly, and those already disabled and out of the work force will tend to be undercounted.

The confusion and inconsistency in evaluating indirect costs is well illustrated by how mortality effects are sometimes handled in CUA. Often, for illnesses that lead to premature death, CUA's include both lost years of life and a measure of lost income. It seems that an individual loses more than his life when he dies, he loses his life and his income. Again, this will systematically bias the resulting CUA against certain groups in the population who choose not to be or are unable to be in the workforce.

From this perspective it is easier, more accurate, and therefore preferable to simply say that an individual suffering premature death loses some number of years of life. The income which would have been earned during those years, all of which he would have been able to use to enhance his life during those years, is part and parcel of the life he would have led. This is the perspective of QOL Definitions 5 and 6 but not that of QOL Definitions 1 through 4.

Adopting this view, one could go a step further and say that "a year of life is a year of life is a year of life," counting everyone equally regardless of their wage rate, whether they were retired, and so forth. Some might buy this argument in favor of using QOL Definition 6 (or 5) but take issue with this treating each individual equally. A case could be made, for instance, that someone with strong labor market skills (due to native intelligence, more education, greater athletic ability, etc.) loses more by dying than someone with less labor market skills. The former individual would have enjoyed a greater quality of life, through his greater income, than the latter individual and this should be taken into account in the analysis. Similarly, some have suggested that a year of life by an elderly individual, even one in robust health, is inherently less enjoyable and provides less utility than a year of life to someone who is younger.

Nonetheless, this monograph proposes that equal weighting of individuals in terms of incorporating indirect costs into a quality of life measure is the appropriate "default" procedure. With this as the default, any differential weighting of individuals by wage or other factors must be advanced with care and done in an explicit and deliberate fashion. The nature of the assumptions and procedures adopted in the differential weighting of individuals would then be well advertised to the consumer of the analysis. This is a far superior situation than measuring indirect costs in monetary terms where, unless very complex corrections are performed there is a default bias against certain groups of individuals.

### *6.3.3. Uncertainty*

A second reason for using QOL Definition 6 concerns the uncertainties in many health interventions. This reason not only supports QOL Definition 6 relative to Definitions 1 through 4 but, in contrast to the first reason, also argues in favor of QOL Definition 6 over Definition 5.

As was mentioned in Chapter 2, an important impediment to CUA for many health interventions is the great amount of uncertainty that is associated with the analysis. As will be discussed in Chapter 11, it is important to recognize and explicitly take account of this uncertainty. This is much easier to do if the relatively certain effects of the health intervention are kept separate from those aspects of the analysis which (1) are more uncertain and (2) tend to be highly correlated (either positively or negatively).

For many health interventions, it may be relatively easy to determine with a fair degree of precision the direct inputs to a health treatment: medication costs, health professional costs, time on the part of the patient, and so forth. Far less certain are the long-run (or even short-run) clinical impacts of a health intervention, let alone their resulting impact on an individual's life, including future work opportunities, future levels of leisure, future years of life, and so forth. Moreover, these impacts on an individual's life tend to vary in the same way. If the health intervention works well all of these things will be more positive than if the health intervention does not work well. These factors, in short, share common sources of uncertainty.

It thus makes sense to group these uncertain quality of life factors together in the denominator of a cost-utility ratio, and to include in its numerator, along with the direct costs of a health intervention, any direct time inputs by the patient. Note that the direct costs of a health intervention and the direct time inputs of the patient are not only both generally relatively certain, but they too tend to vary together. For instance, the number of visits to a doctor to treat an illness may not be known in advance. The direct costs of the treatment are therefore uncertain. However, they will generally vary in the same way as the time inputs by the patient. If the treatment is very effective and requires fewer visits than originally expected, not only are there fewer direct costs but less patient time input as well.

#### *6.3.4. Marginal Versus Nonmarginal Effects*

The third argument in favor of QOL Definition 6 concerns the "lumpiness" issue. The analysis of this chapter has assumed that the components captured in the numerator of a cost-effectiveness ratio and measured as costs are relatively small compared to an individual's overall wealth. This is much more likely to be the case for direct inputs to a health intervention than the indirect ones. It is easy to assume that the \$100 in direct costs for a cancer screening procedure is relatively small compared to the net present value of an individual's current wealth and future income. It is much less easy to make this assumption for the lost years of life and lost future income from not having the test done and developing cancer that would have been averted.

As discussed in Section 4, when a health intervention leads to nonmarginal effects and these are encompassed in the numerator of a cost-utility analysis, complications arise. It can be shown that these complications are avoided when the nonmarginal effects are incorporated in the denominator for cost-utility analysis. Insofar as indirect costs can well be nonmarginal in their impact, CUA is much easier if they are incorporated in the denominator of a cost-utility ratio, as in Definition 6.

#### *6.3.5. Difficulty in Translating Patient Time Input into QOL*

A fourth argument in favor of QOL Definition 6 pertains especially to nonmonetary direct inputs to a health intervention, such as leisure. An individual's subjective evaluation about the monetary value of a unit of leisure, as measured by his willingness to pay for a marginal unit of leisure, may be relatively straightforward. In contrast, it may be much more difficult to derive an accurate measure of the impact of the marginal unit of leisure on overall quality of life.

Looking ahead slightly, this approach to defining the denominator of a cost-utility ratio also has an advantage in terms of workplace health promotion activities. Consider an individual who takes paid time off work to visit a doctor or participate in a worksite health promotion program. This time input does not cost the employee anything and it is quite possibly more enjoyable than working. Rather than try to incorporate this time input in the same quality of life measure, it can now be more routinely accounted for in terms of the cost to the employer of paying the worker during the time involved.

Of the six definitions of quality of life provided, a strong case can be made for adopting Definition 6 for QOL in CUA. According to this definition, the numerator of a cost-effectiveness ratio in CUA should consist only of direct costs and the opportunity cost of time devoted by individuals to the health intervention. The denominator should capture all of the effects of the health intervention on health, social interactions, and the amount and quality of leisure time and work.

#### **6.4. Redefining CUA in Health Care**

In dealing with the type of health interventions considered in Parts I and II of this monograph, in which the only costs are direct medical costs and the only effects are on health, the basic definition of CUA in health care, provided in Chapter 1, remains fully intact. As health interventions involving other attributes are considered, however, the definition of CUA must be modified somewhat.

CUA remains a methodology in which the desirability of a health care intervention relative to the status quo can be gauged by a cost-effectiveness ratio. Now, however, the components of the numerator and denominator can be defined more extensively. The numerator becomes direct costs plus direct inputs by the patient. The denominator is no longer the "aggregate change in health outcomes" that results from a health intervention but "aggregate change in quality of life" that results in a health intervention. Quality of life, in this case, includes all components of overall utility except those that are now captured in the numerator of the cost-effectiveness ratio.

In this sense, the statement made at the beginning of this chapter that quality of life is essentially the same as "overall utility" must now be modified slightly as well. Now quality of life is defined as encompassing  $H$ ,  $S$ ,  $L_s$ ,  $L_u$ , and  $C_i$ , but not  $C_d$  and  $L_d$ .



QALYs AS AN OUTCOME MEASURE FOR CUA

1. INTRODUCTION

Selecting a health intervention often involves making tradeoffs between (1) the intensity and the duration of health impacts, and (2) the morbidity and mortality risks. Thus a health outcome measure in cost-utility analysis (CUA) must be able to account simultaneously for all of these aspects of health states.

Previous chapters have suggested the quality adjusted life year (QALY) as a measure of health outcomes that meets this criterion. Weinstein and Fineberg (1980) define QALYs as "the number of years at full health that would be valued equivalently to the number of years of life as experienced."<sup>51</sup> This allows any health outcome, defined in terms of a profile of an individual's health states over time, to be assigned a QALY value. Thus QALYs, like VNM-utility, merges morbidity and mortality considerations as well as intensity and duration considerations.

While Chapter 9 discussed the possible structures for the quality-of-life measure in the denominator of a CUA, little attention was given to the duration of effect of a health intervention on the overall outcome measure. This chapter deals with durational concerns and addresses the use of QALYs. It discusses the basic framework of QALYs in CUA, conditions under which QALYs are consistent with the theoretical underpinnings of VNM-utility theory, approaches that can be adopted when QALYs cannot be used directly in CUA, and basic principles for eliciting utility judgments regarding health states.

2. USE OF QALYS IN CUA: THE BASIC FRAMEWORK

It is useful to define two benchmarks for quality of life. The first benchmark, with a utility value of 1, is full quality of life. The second, with a utility value of 0, is death.<sup>52</sup> Notationally, these health states are represented as:

$$(10-1a) \quad U(H^*, S^*, L^*, M^*) = 1,$$

$$(10-1b) \quad U(H^o, S^o, L^o, M^o) = 0,$$

where H is health, S is social functioning, L is leisure, and M is material consumption.

Here, the superscript "\*" indicates a high level for the associated variable.  $H^*$  can be defined as "in full health,"  $S^*$  as "fully able to function socially," and  $L^*$  as "fully able to enjoy leisure."<sup>53</sup> It is a little more difficult to apply the notion of "highest possible level" to material consumption M. Perhaps the most analogous definition is to have  $M^*$  represent a level of material consumption

---

<sup>51</sup>Weinstein and Fineberg, op. cit., p. 253.

<sup>52</sup>Recall that assigning the health state "death" a utility value of 0 does not preclude worse health states. Health states worse than death would be assigned negative utility values.

<sup>53</sup>In this formulation, leisure could be divided into home-production leisure, self-consumption leisure, time input associated with a health intervention, and so forth. All that would be required is to define a higher- and lower-level benchmark for each of these types of leisure. The rest of the analysis would proceed exactly as above.

associated with being fully able to function occupationally. The superscript "o" in Equation (10-1b) represents a lower level for each of the associated variables and, as indicated, the quality-of-life state ( $H^o$ ,  $S^o$ ,  $L^o$ ,  $M^o$ ) is defined as the state of death.

When utility is expressed as  $U(H,S,L,M)$ , it represents the utility at a given moment. A subscript "t" can be added to the notation to produce the following notation for the quality of life experienced by individual  $n$  at time  $t$ :

$$(10-2) \quad U_{n,t}(H,S,L,M) = U_{n,t}(H_{n,t}, S_{n,t}, L_{n,t}, M_{n,t}).$$

In this formulation,  $t$  can be thought of for now as a discrete period of time, e.g., a year, a month, or a day.

Consider a health intervention, Health Intervention  $i$ , that is to be compared with the status quo. Let the subscript "o" indicate outcomes under the status quo and the subscript "i" represent outcomes under Health Intervention  $i$ . Additionally, let  $t$  represent a year and let  $t$  be indexed to run from  $t = 0$  (this year) to  $t = T$ .  $T$  can be defined as some distant future year such that there are no differential health impacts between Health Intervention  $i$  and the status quo after time  $T$ .<sup>54</sup>

To measure the change in QALYs associated with Health Intervention  $i$  compared to the status quo, two alternative approaches may be pursued. In the first approach, the cumulative quality of life of individual  $n$  under Health Intervention  $i$  can be computed. Then the cumulative quality of life under the status quo can be computed and the latter figure subtracted from the former. Notationally,

$$(10-3) \quad \Delta QALY_{i,n} = [\sum_t U_{i,n,t}(H,S,L,M)] - [\sum_t U_{o,n,t}(H,S,L,M)],$$

where

$$U_{i,n,t}(H,S,L,M) = U_{i,n,t}(H_{i,n,t}, S_{i,n,t}, L_{i,n,t}, M_{i,n,t}),$$

$$U_{o,n,t}(H,S,L,M) = U_{o,n,t}(H_{o,n,t}, S_{o,n,t}, L_{o,n,t}, M_{o,n,t}),$$

and where  $\Delta QALY_{i,n}$  indicates the improvement in quality of life for person  $n$  under Health Intervention  $i$ , measured in QALYs.

To take a simple example, suppose there are three time periods of relevance,  $t = 0$ ,  $t = 1$ , and  $t = 2$ . Let the utility values for each time period be as follows, where each utility value is calibrated relative to the 1 and 0 benchmark values defined in Equation (10-1):

---

<sup>54</sup>Nothing is changed in the above formulations if  $T$  is set equal to infinity. We choose here for  $T$  to be some finite value simply for expositional convenience.

**Table 11. Individual  $n$ 's Utility Under:**

| Time | Health Intervention $i$ | Status Quo |
|------|-------------------------|------------|
| 0    | .9                      | .4         |
| 1    | .6                      | .2         |
| 2    | .9                      | .7         |

Then,  $\sum_t U_{i,n,t}(H,S,L,M)$  equals  $.9 + .6 + .9$ , or 2.4. Similarly,  $\sum_t U_{o,n,t}(H,S,L,M)$  equals  $.4 + .2 + .7$ , or 1.3. So,  $\Delta QALY_{i,n}$  equals  $2.4 - 1.3$ , or 1.1. Thus, Health Intervention  $i$  improves individual  $n$  by 1.1 QALYs.

An alternative approach is to compute the difference for each year between Health Intervention  $i$  and the status quo and then sum up these differences:

$$(10-4) \Delta QALY_{i,n} = \sum_t [U_{i,n,t}(H,S,L,M) - U_{o,n,t}(H,S,L,M)].$$

It is easy to verify algebraically that Equations (10-3) and (10-4) always give the same result.

Applying Equation (10-4) to the above example, the difference in utilities for each of the three years is  $(.9 - .4)$ ,  $(.6 - .2)$ , and  $(.9 - .7)$ , or .5, .4, and .2. Adding these differences together yields  $.5 + .4 + .2$ , or 1.1, the same value obtained previously.

### 3. EXTENDING THE BASIC FRAMEWORK

#### 3.1. Moving from CUA-I to CUA-II

Having presented the basic framework for using QALYs, we now turn to several extensions. First, the above discussion considered the impact of a health intervention on individual  $n$  alone. Moving from a concern with a given individual to a societal perspective, all that is required is that the impact of the health intervention on each affected party be considered and summed together. The improvement in QALYs from Health Intervention  $i$  then becomes simply:

$$(10-5) \Delta QALY_i = \sum_n \Delta QALY_{i,n},$$

where  $\Delta QALY_{i,n}$  can be defined either by Equation (10-3) or (10-4).

### 3.2. Uncertain Health Impacts

It may well be that the health outcomes associated with Health Intervention  $i$  or the status quo are not known with certainty. In this case, each possible outcome for each time period for each individual must be specified. Then, using the probability of each of these outcomes, the expected utility for each time period must be computed and used in Equations (10-3) and (10-4).

To give a simple example of this type of situation, consider the example from above, modified now in the following table to reflect uncertainty about the outcomes that might be experienced:

**Table 12.** Individual  $n$ 's Utility Under:

| Time | Health Intervention $i$              | Status Quo                           |
|------|--------------------------------------|--------------------------------------|
| 0    | .9 with $p = .5$<br>.4 with $p = .5$ | .4                                   |
| 1    | .6                                   | .2 with $p = .3$<br>.5 with $p = .7$ |
| 2    | .9                                   | .7                                   |

Now there are two uncertainties that must be taken into account in evaluating Health Intervention  $i$ . It is uncertain how helpful Health Intervention  $i$  will be during time period 0. Health Intervention  $i$  may lead to a utility of .9 during time period 0 but, with a 50 percent chance, it yields a utility of only .5. In addition, the situation under the status quo is uncertain during period 1. The individual's utility under the status quo in time period 1 may be as low as .2 but, with a 70-percent chance, it will be .5.

Calculating expected utility for Health Intervention  $i$  for time period 0 yields an expected utility of  $(.5 \times .9) + (.5 \times .4)$ , or .65. Similarly, the expected utility for time period 1 for the status quo is  $(.3 \times .2) + (.7 \times .5)$ , or .41. Then, applying Equation (10-3), the QALY under Health Intervention  $i$  is  $.65 + .6 + .9$ , or 2.15. Similarly, the QALY under the status quo is  $.4 + .41 + .7$ , or 1.51. The value for  $\Sigma QALY_{i,n}$  then becomes  $2.15 - 1.51$ , or .64. As seen in this example, the same approach as used in the prior section applies in calculating QALYs when there is uncertainty about health outcomes, but now expected utilities, instead of the known certain utilities, are used for each time period.

### 3.3. Measuring Time More Finely

The above discussion has taken a unit of time to be a year. If the time interval being used is less than a year, say a month, the same basic approach holds. For each month, the quality of life state of the individual is measured under both the health intervention and the status quo. But now, being in full health for a month does not equal a full QALY, but only 1/12 of a QALY. All that needs to be altered is to redefine Equation (10-1) so that:

$$\begin{aligned} U(H^*, S^*, L^*, M^*) &= 1/12, \\ U(H^o, S^o, L^o, M^o) &= 0. \end{aligned}$$

Similarly, if  $t$  is a day, then Equation (10-1) can be modified as follows:

$$\begin{aligned} U(H^*, S^*, L^*, M^*) &= 1/365, \\ U(H^o, S^o, L^o, M^o) &= 0. \end{aligned}$$

In the extreme, as time is measured in finer and finer intervals,  $t$  can be thought of as a continuous variable. This is also straightforward to handle, although basic calculus notation is required. Now, Equation (10-3) becomes:

$$(10-6) \Delta QALY_{i,n} = [ \int U_{i,n,t}(H,S,L,M)dt ] - [ \int U_{o,n,t}(H,S,L,M)dt ],$$

and Equation (10-4) becomes:

$$(10-7) \Delta QALY_{i,n} = \int [U_{i,n,t}(H,S,L,M) - U_{o,n,t}(H,S,L,M)]dt.$$

### 3.4. More Complex Decision Contexts

So far, this chapter has been concerned with comparing a given health intervention, Health Intervention  $i$ , with the status quo. As Part I demonstrated, the basic mechanics of CUA are such that it is possible to extend the same tools to evaluating more complex decisions, such as choosing among mutually exclusive health interventions and determining the optimal size of a health intervention.

All of these extensions are fully applicable when using QALYs as the outcome measure for CUA. In comparing mutually exclusive health interventions, the value for  $\Delta QALY$  should be computed for each intervention, measured in each case relative to the status quo. This forms the denominator of the cost-utility ratio for that intervention and the mechanics presented in Chapter 3 for choosing among mutually exclusive health interventions can then be applied directly. Similarly, for health interventions that can vary in scale, each scale size,  $C$ , should have a value for  $\Delta QALY(C)$  computed and the mechanics presented in Chapter 4 can then be applied directly to determine the cost-effective scale for the intervention.

### 3.5. When Can Non-Health Attributes Be Ignored?

In Chapter 9, the notion of additive separability of utility functions was introduced. To use material consumption,  $M$ , as an example, the multiattribute utility function for an individual's quality of life is additively separable with respect to material consumption if Equation (10-2) can be written as (10-2):

$$U_{n,t}(H,S,L,M) = U1_{n,t}(H_{n,t}, S_{n,t}, L_{n,t}) + U2(M_{n,t}),$$

where  $U1(.)$  is some function of  $H$ ,  $S$ , and  $L$  and  $U2(.)$  is some function of  $M$ . Then,

$$(10-8) \quad U_{i,n,t}(H,S,L,M) - U_{o,n,t}(H,S,L,M) =$$

$$U1_{i,n,t}(H_{i,n,t}, S_{i,n,t}, L_{i,n,t}) + U2(M_{i,n,t}) - U1_{o,n,t}(H_{o,n,t}, S_{o,n,t}, L_{o,n,t}) + U2(M_{o,n,t}) = \\ [U1_{i,n,t}(H_{i,n,t}, S_{i,n,t}, L_{i,n,t}) - U1_{o,n,t}(H_{o,n,t}, S_{o,n,t}, L_{o,n,t})] + [U2(M_{i,n,t}) - U2(M_{o,n,t})].$$

Now suppose that Health Intervention  $i$  does not affect material consumption, so that  $M_{i,n,t}$  equals  $M_{o,n,t}$ . Then, Equation (10-8) becomes:

$$U_{i,n,t}(H,S,L,M) - U_{o,n,t}(H,S,L,M) =$$

$$U1_{i,n,t}(H_{i,n,t}, S_{i,n,t}, L_{i,n,t}) - U1_{o,n,t}(H_{o,n,t}, S_{o,n,t}, L_{o,n,t}).$$

Note that the component of overall utility having to do with material consumption,  $U2$ , is now completely removed from the equation for calculating QALYs. In this case, nothing to do with  $M$  plays any role in calculating QALYs.

This result holds for other components of overall utility besides  $M$ . As long as that component is additively separable from the other components in the quality of life utility specification, and if the health intervention does not affect that component, then it need not be taken into account in measuring QALYs associated with the health intervention.

In the extreme, when overall utility is additively separable in all its components and a health intervention is "simple" in the sense that it affects only health and not an individual's material consumption, leisure, or social functioning, then Equation (10-4) becomes:

$$(10-9) \quad \Delta QALY_{i,n} = \sum_t [U(H_{i,n,t}) - U(H_{o,n,t})].$$

That is, QALYs can be computed simply by considering the difference in health between having the health intervention and not having it. Other features of quality of life need not be considered. In this case, any distinctions between QALYs and wellness years then become moot as both notions reduce to the same concept.

#### 4. QALYS AND VNM UTILITY

It has been presumed throughout this chapter that the use of QALYs is compatible with VNM-utility theory and therefore appropriate for CUA. But is this really so? The answer turns out to be a conditional yes. The QALY is a proper scale for measuring health outcomes in CUA only if individuals' preferences regarding health states satisfy certain properties. While proving that these properties are necessary for QALYs to be consistent with VNM-utility theory and, therefore, with CUA, is beyond the scope of this work, it is useful to indicate these properties in order to understand the strengths and limitations of QALYs as an outcome measure for CUA.<sup>55</sup>

---

<sup>55</sup>These assumptions are developed more formally in J.S. Pliskin, D.S. Shepard, and M.C. Weinstein, "Utility Functions for Life Years and Health Status: Theory, Assessment, and Application," *Operations Research*, 28:206-24, 1980. See also, Weinstein and Fineberg (1980), p. 218-19.

The first property necessary for QALYs to be appropriate for CUA is *utility independence* between length of life and quality of life. This means that tradeoffs between length of life are not affected by the quality of life experienced by the individual. Similarly, tradeoffs among levels of morbidity are not affected by the length of life that the individual will live.

These properties can be illustrated using the examples of Weinstein and Fineberg (1980). Consider first tradeoffs between length of life and quality of life while alive. Suppose an individual is indifferent about living either 14 years with mild angina and playing a lottery in which there is a .5 chance of living 10 years with mild angina or a .5 chance of living 20 years with mild angina. Note that in this choice, all the health outcomes are such that while alive the individual experiences mild angina. The tradeoff faced involves length of life and not quality of life while alive.

Utility independence implies that the individual will have the same preferences over length of life if, in making a choice among these health outcomes, "mild angina" is replaced by "no angina" for each health state. That is, the individual must be indifferent between living 14 years with no angina versus a lottery in which there is a .5 chance of living 10 years with no angina and a .5 chance of living 20 years with no angina.

Similarly, under utility independence, tradeoffs between morbidity levels are not affected by length of life considerations. For instance, suppose an individual is indifferent about either living 20 years with mild angina or a lottery in which there is a .5 chance of living 20 years with no angina and a .5 chance of living 20 years with severe angina. Note that, for all of the health states in this choice situation, the individual lives 20 years. The tradeoff here is between levels of angina, not length of life. Under utility independence, the individual must display the same tradeoffs between levels of angina if the length of life associated with each outcome is changed to, say, 10 years instead of 20. That is, the individual must be indifferent between living 10 years with mild angina and a lottery in which there is a .5 chance of living 10 years with no angina and a .5 chance of living 10 years with severe angina.

Finally, utility independence assumes that the timing or ordering of health states does not affect preferences. For instance, the utility (or disutility) of having a toothache next week is not affected by whether one is going to have another toothache in 6 months. In a similar vein, utility independence implies that an individual will be indifferent about either breaking his leg this month and having a toothache next year or having a toothache this month and breaking his leg next year.

The second preference property required for QALYs to be a valid outcome measure for CUA is the *proportional trade-off property*. To illustrate this property, suppose an individual is indifferent about either 32 years of life with no angina or 40 years of life with angina. That is, he is willing to give up one-fifth of his 40 years to be without angina. Then, the individual must also be indifferent between 24 years of life with no angina and 30 years of life with angina. That is, he should again be willing to give up one-fifth of his years of life to be without angina.

The final preference property required for QALY to be used in CUA is *risk neutrality* with respect to years of life. Under risk neutrality, an individual should be indifferent about either 5 years of life in full health and a .5 chance of living 0 years longer or a .5 chance of living 10 years in full health. More generally, he must be indifferent between a gamble over length of life and having with certainty a length of life equal to the expected value of the gamble.

## 5. WHAT IF INDIVIDUALS ARE "RISK AVERSE" OVER LENGTH OF LIFE?

The above three preference properties should not be confused with the basic axioms that underlie VNM-utility theory, discussed in Chapter 6. There is nothing "irrational" about an individual's preferences if they do not follow these properties. In fact, it is clear that in many cases, these properties are only approximations at best.

For instance, several empirical analyses indicate that many individuals are not risk neutral towards length of life. An example is provided by McNeil et al. (1981). Holding quality of life constant, individuals were presented with the following choice: a 50:50 gamble in which death is either immediate or occurs after 25 years, and the certainty of living  $X$  years. Individuals were asked the value of  $X$  for which they were indifferent about the two choices. Individuals who are risk neutral with respect to length of life will assign a value to  $X$  of 12.5. McNeil et al. found, however, that the average value of  $X$  for the individuals examined in their study was 8 years. That is, the average individual was indifferent with respect to having a 50:50 change of immediate death or living 25 years as compared to knowing with certainty that he will live 8 years.

In facing gambles for money, an individual is said to be risk averse if he or she prefers the expected value of a gamble to taking the gamble. Extending this notion to gambles over length of life, the average individual in the McNeil et al. analysis can be said to be "risk averse" over length of life.<sup>56</sup> The finding of risk aversion with respect to length of life is not surprising. Most analyses of risk attitudes to money find the same sort of risk aversion.<sup>57</sup> Indeed, risk aversion with respect to money is considered so typical that it is often thought (mistakenly) to be one of the axioms of rational choice. If most people's attitudes toward length of life are similarly risk averse in most situations, then the risk neutrality assumption, necessary to justify using QALYs in the manner presented in the prior section, will be violated.

When the risk neutrality preference property is not valid and is violated sufficiently that it does not seem useful as an approximation, CUA need not be abandoned as a method of analysis. Instead, a given health outcome—including both intensity and duration information—must be defined and all other health outcomes can be measured relative to it. For instance, consider the health outcome "living one more year in good health." Call this measure a "google," (a made-up name). Other health outcomes can be measured in "google-utility units." With all health outcome states measured in terms of google-utility units, all of the normal CUA procedures can still be applied.<sup>58</sup>

To illustrate, suppose an individual is risk averse with respect to length of life (i.e., the risk neutrality preference assumption is violated) and is indifferent about accepting either one year of additional life

---

<sup>56</sup>To see this, note that the expected value of the McNeil et al. gamble is 12.5 years, which is preferred by the average subject in the study to the gamble itself (which is rated as equivalent to having 8 years of life with certainty).

<sup>57</sup>In the context of gambles over money, risk aversion is sometimes explained by a diminishing marginal utility argument. Diminishing marginal utility means that, while additional money is always beneficial, the incremental level of benefit goes down the more money one already has; in short, a dollar to a rich person means less than a dollar to a poor person.

<sup>58</sup>One could even call a "google" a quality adjusted life year and be back to the original terminology, although care would have to be taken to make clear that QALYs now cannot be added together before being normalized through the above type of procedure.

at full health (with no uncertainty) followed by death and a lottery giving a .5 chance of immediate death or a .5 chance of 3 years of additional life at full health. Then, 3 years of life at full health would be assigned a value of 2 googles.<sup>59</sup> This differs from the QALY value of 3 that would be assigned if the individual were risk neutral.

Some simple mathematical functions can play a useful role in dealing with preferences that are risk averse towards length of life. For instance, consider the functional form:

$$(10-10a) \quad u(x) = a(1 - e^{-cx}),$$

$$(10-10b) \quad a = 1/[1 - e^{-c}],$$

where  $x$  is length of life,  $e$  denotes the number equal to approximately 2.718 (and defined more precisely to be the number such that  $\ln(e) = 1$  where  $\ln(e)$  stands for natural logarithm, and  $u(x)$  is the utility (measured in, say, google-units) of a length of life of  $x$  years. This is a commonly used functional form in utility theory. If  $c$  is 0, the individual is risk neutral. If  $c$  is positive, the individual is risk averse. The greater  $c$ , the more risk averse the individual. The parameter " $a$ " is a normalization parameter and its value indicated in Equation (10-10b) is chosen so that  $u(0) = 0$  and  $u(1) = 1$ . That is, Equation system (10-10) is set up so that one year of life is valued at one utility unit and zero years of life is valued at zero utility units.

While quite flexible in capturing a wide range of risk preferences in a very simple way, the utility function in Equation (10-10) can be shown to have some features that may make it nonrepresentative of individual preferences.<sup>60</sup> An alternative, and also quite tractable, class of functions for capturing risk aversion is:<sup>61</sup>

$$(10-11) \quad u(x) = x^{1-c}, \quad c < 1,$$

where, again, it is easily verified that  $u(0) = 0$  and  $u(1) = 1$ . The case of risk neutrality corresponds to  $c = 0$ . The greater the value of  $c$ , the more risk averse the individual is with respect to length of life.

---

<sup>59</sup>To see this, note that if three years of additional life is assigned 2 google-utility units, then the lottery will have an expected utility of  $(.5 \times 0 \text{ googles}) + (.5 \times 2 \text{ googles})$ , or 1 google. Then, by VNM-utility theory the individual will indeed be indifferent between one additional year of life at full health, with an expected utility value of 1 google, and the above lottery, with an expected utility value of 1 google.

<sup>60</sup>It can be shown that  $u(x)$  is of the form indicated in Equation (11-10) if and only if any fixed amount of additional years of life can be added to all outcomes in a choice among gambles without affecting preferences (see Keeney and Raiffa, p. 167). For instance, suppose an individual is indifferent about either a 50:50 gamble at living 0 years or 25 years or knowing he will live for 8 years (as was typical in the McNeil et al. study). Then for  $u(x)$  to be of the form indicated in Equation (11-10), an individual must be indifferent about either a 50:50 gamble at living 10 years or 35 years, or knowing for sure he will live for 18 years, where 10 years has been added in this case to all outcomes of the gamble.

<sup>61</sup>It can be shown that utility functions such as in Equation (11-11) display "constant proportional risk aversion." In the context of the McNeil et al. example, this means that all outcomes of the gambles can be multiplied by a fixed quantity without changing preferences over the gambles. For instance, the individual who is indifferent between 8 years of life with certainty and a 50:50 gamble over 0 or 25 years of life will also be indifferent between 16 years of life with certainty and a 50:50 gamble over 0 or 50 years of life. See, John W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, 32(1-2):122-36, 1964.

For most practical purposes, Equations (10-10) and (10-11) should provide a sufficient breadth of functional forms to use in capturing risk aversion over length of life for CUA. The reader is referred to Keeney and Raiffa (1976), Chapter 4, for a discussion on how preferences over length of life can be elicited and, thereby, how a numerical value for the parameter  $c$  can be determined for these equations.

## **6. SOME OTHER CAVEATS WITH QALYS**

### **6.1. Quality of Life Versus Health**

Apart from possible violation of the risk neutrality assumption, several other caveats pertain to use of QALYs in CUA. The first follows from the discussion of Chapter 9. Seldom does the CUA literature explicitly define the notion of quality of life in defining QALYs. As in the Weinstein and Fineberg definition, a full QALY is typically defined as a year in good health. It is important to modify this definition when the multiattribute nature of quality of life is recognized. As defined here, a full QALY should be considered a year at full health, social functioning, and work capability. As indicated above, only when a health intervention is "simple"—i.e., only health is affected directly or indirectly by the intervention—and only when health is additively separable from other components of quality of life, can this distinction between health and quality of life be ignored.

### **6.2. Advanced Knowledge and Anxiety About Future Health Outcomes**

Another issue in defining QALYs concerns the timing of an individual knowing about future health outcomes. Suppose an individual is told that in 5 years he will experience a serious illness and there is nothing he can do about it between now and then. Once the illness is present, he can immediately have surgery. If successful, he will live 5 additional years. If unsuccessful, he will die on the operating table. Without the surgery he will live only 6 months. The surgery has success probability of .5. The individual may well indicate that when the time comes, he will prefer the surgery.

Contrast this situation with one in which the same individual is told that he *now* has the serious illness and he faces immediate surgery, with the same odds as above. The individual may now prefer the option of having 6 months of added life with certainty if he feels his affairs are not in order and it is essential that he bring them into order during his remaining time. As this example indicates, the surprise or suddenness of a health outcome can have a substantial influence on an individual's utility judgements for health states, although this is not generally taken into account in the existing literature.

In defining a QALY, what base-case assumption should be invoked about the level of knowledge an individual has about his fate? Any level could be chosen in principle and departures from that level of knowledge calibrated accordingly.<sup>62</sup> Still, to make QALYs operational, some base-case assumption about the level of knowledge should be provided. It seems reasonable to define a QALY for CUA in terms of a year of life for which an individual has sufficient time to "put his affairs in order." Extended to morbidity, the base case is that an individual has enough warning to plan for the illness. For instance, in facing minor surgery, the individual is able to find someone to stay with the

---

<sup>62</sup>Two extremes might be considered. One extreme is full knowledge: "you will die in 1 year and can take whatever actions are appropriate between now and then with this information in mind." At the other extreme one might consider hypnotizing an individual, telling him about prospective health outcomes and eliciting his preferences. Upon coming out of hypnosis the individual will be as unaware about future health states as before the elicitation.

kids, to rearrange important business appointments, and so forth. If a given health intervention concerns outcomes for which such preparation time is lacking, this should be taken into account in measuring the disutility of the associated outcome, calibrated relative to the proposed base-case.

A related issue concerns anxiety about upcoming health states. In choosing between having a root canal today or putting it off for a year, an individual may want to put it off: why go through pain now if it can be postponed? On the other hand, knowing that a root canal will be needed in the future may lead to apprehension and anxiety. The individual therefore may choose to have it done today.

The possible presence of such anxieties is another reason that it is important to specify clearly the certainty that an individual has about an upcoming health outcome. In keeping with the above discussion, it would seem reasonable to assume that a QALY is defined such that an individual knows the specific timing of the health outcomes. If a health intervention leads to a health outcome for which the specific timing is less certain, and if this causes less anxiety for the individual than knowing the specific timing with certainty, then this should also be taken into account in valuing the respective health outcome.

### **6.3. Impact of One Person's Health on Others**

A different kind of limitation to QALYs as they typically are employed is their focus on the utility of the individual directly affected by the health intervention. In many instances, however, there are significant impacts on the quality of life of friends and family when an individual becomes ill or dies. Some efforts to account for "spillover" effects on others can be made directly. For instance, one person's illness can impose time costs on other family members in caring for him, taking him to the doctor, and so forth. The same methods for attaching a monetary value to the time input of the ill individual can be used to attach a monetary value to the time input of family members.

Typically, however, the biggest spillovers from an individual's illness are more intangible, and currently CUA is not able to account for them very well. Measuring the grief a family feels when a member dies, for instance, is beyond the state of the art in CUA and is a general limitation to CUA in its present state of progress. Better measurement of these impacts is an important area for future CUA research over the upcoming decade.

### **6.4. Are All Years the Same?**

A related limitation of QALYs is its assumption that all years are the same, equal, and interchangeable. Some have argued that there is a different level of tragedy to losing a year of life depending on the age of an individual. Some empirical analyses, for instance, indicate that many people find the death of a very young child less tragic than that of a thirty-year-old father. In part, this is undoubtedly due to the impact of the father's death on his family. Again, the current state of the art in CUA does not easily address these considerations.

More generally, even apart from the impact of one person's illness on others, using QALYs in CUA assumes comparability of health and quality of life states across individuals. A year of life at a given level of health (and social functioning, leisure, and material consumption) is viewed as comparable

across individuals.<sup>63</sup> This is consistent with the interpersonal utility comparison assumption presented in Chapter 7, but runs counter to the basic presumption in most of neoclassical economics that individuals' welfares cannot be compared in the sense of indicating whether one person enjoys something more than another. Indeed, much of the edifice of neoclassical economics has been built by purposefully rejecting interpersonal comparability of utility across individuals.<sup>64</sup> In this regard, CUA departs sharply from the traditional neoclassical economic perspective and, as discussed in Chapter 7, is much more "utilitarian" in its orientation.

## 6.5. A Caveat to the Caveats

The concerns raised in Sections 4 through 6 are legitimate caveats to the use of QALYs in their simplest form in CUA and are recognized or taken into account in the existing literature too infrequently. At the same time, their importance should not be exaggerated. Often the utility assumptions presented in Section 4 are reasonable approximations of individuals' preferences in regard to health outcomes, and the concerns raised in this section may be of only minor importance for the health interventions considered in a particular study.

It should also be emphasized that while some of these concerns are of sufficient importance that they cannot be easily ignored without doing substantial damage to the analysis, the principles behind conducting CUA still remain intact. Only now, various measurement modifications need to be done. These are quite feasible in principle, even if they are seldom if ever done in practice.

CUA has evolved substantially over the past dozen years. It might be hoped that over the next dozen years it will continue to evolve at a comparable pace. If so, many of the types of concerns raised in Sections 4 through 6 will be more widely acknowledged and, where relevant, taken into account as a matter of course.

## 7. ELICITING QALYS

### 7.1. Introduction

Sections 4 through 6 presented some qualifications to the use of QALYs in CUA. However, it must be noted that there is an even more basic practical concern with using QALYs: there are relatively few empirical elicitations of utility values for health states (let alone quality-of-life states). Unfortunately there is no master list of all known health and quality-of-life states to which one can refer in assigning utility or QALY values.

---

<sup>63</sup>This does not mean that it is necessary in CUA to assume that each person experiences a given type of illness exactly the same. Someone with trigeminal neuralgia, a form of facial pain, may experience a toothache much more severely than someone else who has the same type of tooth infection. To the degree to which these differences in sensitivities can be stated and measured, then different individuals can experience the "same" health outcome in different ways and this can be fully incorporated into the framework of CUA. Nonetheless, different individuals with the same underlying characteristics *are* presumed in CUA to experience a given type of health outcome in the same way.

<sup>64</sup>The very notion of economic efficiency and its reliance on the Pareto principle (which states "one state of society is preferred to another if some individuals consider themselves to be better off in the preferred state and no one considers themselves worse off") has been carefully crafted to avoid comparisons of utility across individuals.

In conducting a CUA it is often necessary instead for the analyst to determine utilities and QALY values for the health and quality of life outcomes that arise in the context of the analysis. How can this be done? The most common approach by far is to elicit such judgments from individuals directly. In doing so, individuals are asked, using one methodology or another, the level of utility they place on certain health outcomes. Their responses form the basis for measuring utility levels.

A detailed presentation of how to elicit utility preferences concerning health states is beyond the scope of this monograph. However, it is important to present some of the general issues that arise in such elicitations.

## **7.2. Elicitation Versus Revealed Preference Approaches**

To begin with terminology, the determination of utilities by eliciting judgments is sometimes called "contingent valuation." This term, supplied by CBA, is used to denote circumstances in which an individual's willingness to pay for a good (typically an intangible good such as health) is elicited from the individual. This is usually done by presenting a hypothetical situation to the individual concerning the good in question and inferring from the answer how much the individual would be willing to pay if there were actually a market for the good.

In CUA, the interest is not with the willingness to pay, and there is no need to create a market setting in making choices. Despite these differences, the issues of contingent valuation methodology in CBA overlap substantially with those that arise in developing health state utilities.

The approach of contingent valuation, and of utility elicitation more generally, can be contrasted with that of "revealed preference." Revealed preference uses actual choices that the parties involved have made, most typically choices in a market setting, to infer underlying preference patterns. When feasible, revealed preference is often preferred by economists.

They argue that hypothetical responses to questions about how much something is worth are unlikely to be reliable. For instance, consider asking about the importance of clean air in an attempt to determine the value placed on the health improvements that result from reducing pollution. Asked this hypothetical question, an individual could easily respond with a very high number. Such a response may reflect distaste for pollution *and* the fact that there is no need to pay this amount to get the pollution reduced. However, in observing actions the same individual takes that involve trade offs between money and pollution (e.g., how much an individual is willing to pay to have a house that is located where there is less pollution), the individual may really be willing to spend only a small fraction of the stated amount. The same concern arises in eliciting utilities. The problem is that, for many health interventions, it is very difficult or impossible to use a revealed preference approach. It may be hard even to determine what observation of market-based activities would show the value placed on conditions like not being on kidney dialysis or not being depressed.<sup>65</sup>

However, just because there are potential difficulties in implementing a utility assessment does not mean that such assessments are inherently unreliable. Over the past dozen years there have been

---

<sup>65</sup>This is not to say that revealed preference methods are always impossible nor that they should not be pursued when feasible. For instance, in assessing the disutility of side effects for some medication, it would be valuable to rely not only on stated preferences about the disutilities experienced but also behaviors that are associated with the disutilities—such as refusing to take the medication, dropping out of an experiment in which the medication is being tested, and so forth.

tremendous advances in techniques to elicit utility judgments as well as research into pitfalls that might make such judgments less reliable. This work spans areas such as multiattribute-utility theory, behavioral-decision theory, health policy, and economics proper. Some of the key themes of this work are explored in the following sections.

### 7.3. The Simple Theory of Holistic-Utility Elicitation

In eliciting utilities, one approach is to describe a health state in its entirety and ask the individual to make a judgment about it. This approach can be termed "holistic" health-state assessment and it is the most common approach in the CUA literature. The basic principles of holistic-utility elicitation are straightforward. There are two basic assessment methods. Assuming that the three preference assumptions presented in Section 4 are satisfied, either elicitation method should lead to the same set of values for health states.

One approach is based on the lottery method, similar to that outlined in Chapter 6. For instance, consider an individual on dialysis who faces the possibility of a kidney transplant. Assume, for the sake of argument, that the two outcomes of the kidney transplant operation are death and complete wellness. Assume that either without the operation or after a successful operation, the individual can be expected to live 25 years. Then, the decision about whether to have the operation can be thought of as a lottery with a probability of  $p$  of complete wellness for 25 years and a probability of  $1-p$  of immediate death versus accepting life on dialysis. At what level of  $p$  is the individual indifferent about either having the operation or not? Let this value be denoted as " $p$ ." Then the QALY value of living for 25 years with dialysis is simply  $(p \times 25)$  QALYs, or the QALY value of living one year on dialysis is  $p$  QALYs.

A second way to calibrate health states using QALYs is the "time-tradeoff method." Consider again the same dialysis example. Now suppose the individual is asked the following question: instead of living 25 years with dialysis, you can live  $X$  years at full health instead. At what value of  $X$  is the individual indifferent about the two choices? Let this value be denoted by " $X$ ." Then 25 years with dialysis is equivalent to " $X$ " QALYs, or the QALY value of living one year on dialysis is " $X$ "/25.

### 7.4. Whose Utilities Should be Assessed?

While the principles in measuring utilities are straightforward, a variety of practical issues arise. For example, whose preferences should be elicited? One possibility is to elicit the preferences of those who will be experiencing the health intervention in question. This would seem the most appropriate group (or individual) from whom to elicit preferences. It may not be practical to elicit utilities from these individuals, however. As an alternative, the utilities of individuals who are representative of those who will be affected by the program may be elicited (e.g., Churchill, Morgan and Torrance, 1984).

Even here, elicitation difficulties arise if the person whose preferences are being elicited is not capable of making what might be considered informed judgments. As Fischhoff and Firby (1990) indicate, the key goal for utility elicitation should be that the health and quality-of-life scenario being

considered is understood by the elicitee and evaluated properly.<sup>66</sup> This requires that elicitees are "fully informed, uncoerced, and able to evaluate their own best interests."<sup>67</sup>

Ensuring that judgments come from individuals who are fully and properly informed is particularly challenging in determining preferences for health outcomes. One reason for this is that health states are often novel. The elicitee may have had limited direct experience with the health state. He may seldom have had to think hard about the overall desirability of a health state. Unlike elicitation of preferences over tangible goods, health states are not traded in the market and are not bought and sold and there is no analogue to the trial and error provided in a market in determining if one situation is preferred to another—in this case, whether one health state is preferred to another.

For instance, it may be hard for respondents to give meaningful preferences about home dialysis if they have never had home dialysis or known anyone who has. Similarly, the individual involved may not be judged capable of making informed decisions; e.g., he or she may be currently depressed or in great pain. In health care, habits and addictions, both physical and psychological, may greatly influence individuals' judgments concerning health outcomes.

One approach to the informed judgment issue is to turn to physicians or other health professionals to make the judgments involved (e.g., Weinstein, 1981; Torrance et al., 1973; and Pliskin et al., 1980). This widely adopted practice has much to recommend it. Apart from the familiarity of physicians or other health professionals with the disease, the elicitation of preferences about different health outcomes may be much simpler to execute. To an individual who has never been depressed, a great deal of description may be involved in portraying the nature of depression, the impact on family and friends, and so forth. Such detailed description may not be required for a psychiatrist who works frequently with depressed patients. All that may be required is the score of the patient on some clinical depression measure and a few bits of information about length of the depressive episode and so forth.

Despite the advantages, this approach carries with it several potential drawbacks as well. While physicians may be more informed from a clinical perspective about the health states under consideration, the same clinical perspective may make them insensitive to certain quality-of-life considerations. Similarly, physicians and other health professionals often have personal characteristics that are different from those of the people experiencing the health interventions. They may differ, for example, in sex, age, income, health habits, family composition, or educational level, all of which may influence utility judgments.

## 7.5. Describing Health States

In conducting the elicitation, a judgment must be made about how the health state should be described. In part, this is a function of the identity of the elicitee. If the elicitees are not health

---

<sup>66</sup>Fischhoff and Firby provide a survey of key issues in contingent valuation methodology and proposed a conceptual framework for thinking about the adequacy of contingent valuation elicitation instruments. While their work addresses elicitee's expressions of willingness to pay (and focuses on the value of atmospheric visibility rather than health), it provides valuable guidance for elicitation of health utilities as well.

<sup>67</sup>Baruch Fischhoff and Lita Firby, *Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility*, *Journal of Risk and Uncertainty*, 1 (1):147–84, 1988.

professionals, then the description of the health states as presented to them should be in functional rather than clinical terms.

To the degree to which a health intervention affects social functioning, work functioning, and ability to enjoy leisure, these impacts (and their associated durations) should also be clearly stated. A health professional experienced in a given area may be able to infer these impacts from a clinical description of the health state itself, but a more general audience will need to have these impacts made more explicit. Unfortunately, few CUA's have followed through on this basic requirement for elicitation. Often the work, social, and leisure impacts of an illness are very vague in elicitation scenarios.

There are dangers inherent in either a too detailed or a too terse description of health states. If the health state is relatively novel, it is important that it be described in some detail. Fischhoff and Firby write

There are...limits to people's imaginations, especially with more novel goods and settings...In general, the more novel a transaction, the more of its details will need to be explained and the more difficult it will be to ensure that those details are understood.<sup>68</sup>

Thus, at least for members of the general public, a rather extensive presentation of information may be necessary. On the other hand, Fischhoff and Firby caution that:

Simply telling people everything provides no guarantee that they have understood it all. That strategy might even impede understanding if attention to critical features...is diverted by a deluge of details about features that could have "gone without saying."<sup>69</sup>

A related concern is the question of which items need not be specified. Inevitably, many aspects of the health state, or the context in which the health state occurs, are not mentioned explicitly in the health-state description and are left to be implicitly assumed by the elicitee. An explicit statement of all contingencies and ancillary circumstances generally will be of impractical length and will lead to concerns with "information overload." On the other hand, it is important that whatever the elicitee infers is consistent with the intentions of the elicitor.

Empirically, there have been a wide range of practices in terms of the detail of the health-state descriptions. As indicated in Torrance (1986), some elicitations have used health states described by only a few key phrases, e.g., Patrick et al. (1973) and Kaplan et al. (1976). Others have used "high fidelity" descriptions including video and audio tapes, e.g., Cadman and Goldsmith, 1982;<sup>70</sup> McNeil et al., 1981; and Boyd et al., 1982. Still others have been in the middle, presenting a few paragraphs of description for each health state e.g., Torrance et al., 1973; Sackett and Torrance, 1978.

---

<sup>68</sup>Fischhoff and Firby, p. 9.

<sup>69</sup>Fischhoff and Firby, p. 9.

<sup>70</sup>Cited in George W. Torrance, *Measurement of Health State Utilities for Economic Appraisal: A Review*, *Journal of Health Economics* 5(1), 1986.

Clearly, each elicitation requires a tradeoff between ensuring that the elicitee understands the key aspects of the health state under consideration and protecting against the elicitee being overwhelmed by details. Given the present state of the art, this tradeoff is as much an art as a science.

#### **7.6. Concerns Raised By Behavioral-Decision Theory**

As if the concerns listed were not sufficient to indicate the challenges in eliciting utility judgments, work over the past decade in behavioral-decision theory points to several additional complications. Much work in psychology and behavioral decision theory has emphasized the important differences in response that can be elicited depending on how a question is phrased (Tversky and Kahneman, 1982), where it is placed with respect to other questions (Lau and Kinder, 1984), and even the mood of the respondent (Johnson and Tversky, 1984).

This research also suggests that individuals have particular difficulty in accurately calibrating judgments involving small or large probabilities (Tversky and Kahneman, 1974; Kahneman and Tversky, 1979), and that individuals are especially poor in processing information about low-probability high-loss events. As a result, this literature implies that it is best to avoid judgments over health states that involve very large (.999) or small (.001) probabilities. This is sometimes difficult because in health policy, many outcomes are of this low-probability high-loss nature.

Another finding of behavioral decision theory is that individuals often assess a state of the world not in terms of its absolute characteristics but in terms of the extent to which the new state represents a change from a *reference level*. This phenomenon has been noted in a wide range of contexts. To our knowledge, there has not been empirical work examining the impact of reference level effects on health-state utility elicitation. Based on work in other fields, however, special attention should be paid to specifying the reference level by which health states are to be measured. This also means that the elicitor must determine the reference level which is appropriate. This may not always be easy to do.<sup>71</sup>

Not only do individuals seem to gauge states of the world relative to reference levels but, more particularly, they seem to evaluate "gains" from a reference level differently from "losses." A given loss from a reference level is often considered to be more meaningful than a comparable gain. Frequently, a given health state, or change in health state, can be cast as a gain or a loss. The elicitor should be aware of whether a health state is seen as a gain or a loss from a reference level and the impact of this perception on responses.

In all of these areas, there has been little systematic research applied to CUA and utility elicitation of health states. Scanning the current state of the art, Torrance (1986) concludes that

the best current advice in measuring utilities on the general public is probably to use abbreviated descriptions to avoid cognitive overload, to supplement those with prior more detailed explanations of the key phrases used in the abbreviated descriptions,

---

<sup>71</sup>The influence of reference levels on utility judgments raises the possibility that a given health state does not have a single utility in and of itself but different utilities depending on the reference level. This, in turn, raised difficult issues about the change in utility judgments. To the degree to which utility judgments are excessively changeable and subject to manipulation by seemingly arbitrary rewordings of the health state being described, the whole notion of preferences over health states is called into question.

and to avoid the framing bias by wording the question in a balanced (positive and negative) manner.<sup>72</sup>

Torrance's recommendation is well considered. These issues are important for future research in CUA.

---

<sup>72</sup>Torrance, 1986, p. 13.

## **CHAPTER 11**

### **ADDITIONAL CONSIDERATIONS: UNCERTAINTY AND DISCOUNTING**

#### **1. INTRODUCTION**

This chapter provides an overview of additional practical issues that arise in performing a CUA. Some fall into the category of uncertainties that practitioners may encounter; others are issues involved in discounting costs and health outcomes over a period of time. While the discussion is somewhat abbreviated in the interest of space, it should allow the reader to appreciate key areas of debate and controversy in dealing with these issues, as well as to develop an understanding of common practice.

#### **2. DEALING WITH UNCERTAINTY IN CUA**

##### **2.1 Introduction**

CUAs usually involve many uncertainties. Often, for example, the precise clinical efficacy of a health intervention is uncertain. Or it may not be known exactly how clinical effects translate into impacts on quality of life.

It is important that a CUA accurately and explicitly indicate the confidence or lack of confidence that a reader should place on specific numerical conclusions. This is especially true because users of CUAs may tend to treat the conclusion as a magic number, although it is actually only a best guess about a health intervention's cost per QALY. Moreover, users may completely disregard warnings about the tentativeness of the conclusion, particularly in the context of partisan debate within the policy process. In this arena, nuances about uncertainties are often ignored in an attempt to make one party's favored health intervention appear as desirable as possible. In the effort to make informed decisions about resource allocation in health care, CUA can be a clear disservice if artificial precision is attached to conclusions that, in fact, are highly uncertain.

Indeed, the uncertainties encountered in CUA constitute one of the most common objections to its use. It is often said that a fancy mathematical approach such as CUA is all well and good for "ideal" data but that in practice, sophisticated modeling is of little use in dealing with data that cannot support such fine manipulation. From this perspective, the uncertainties involved swamp the rest of the analysis; one is left with "garbage-in garbage-out." In short, there is often suspicion that CUA adds artificial precision to what simply is not precise.

These are legitimate concerns. Lave and Lave (1978), for instance, pointed out large gaps in the information needed to do a complete economic analysis of most disease prevention and health promotion interventions—gaps not only in knowledge about economic costs but in knowledge about clinical effectiveness and about how clinical impacts affect quality of life. Now, almost 15 years later, there is more information for some types of prevention and health promotion interventions, but for many the situation is much as it was then. The workshop by Russell et al. (1987) is a good case in point, as several of the key prevention areas examined were deemed to have insufficient data to allow for formal cost-effectiveness analysis.

At the same time, the presence of uncertainties, even quite substantial uncertainties, need not preclude CUA. Moreover, resource allocation decisions still must be made for many health interventions,

regardless of the quality of the existing data. Refraining from systematic analysis in areas in which there is uncertainty does not preclude the need to make resource allocation decisions in these areas. Therefore, it is desirable to push the analysis as far as possible, even in the presence of uncertainties.

## **2.2. An Example of the Potential Impact of Uncertainty on CUA**

Sometimes a little uncertainty can have a large impact on a CUA. For example, consider a screening procedure that helps identify individuals who might develop disease Y. Disease Y costs \$10,000 to cure and results in a health loss of .05 QALY. For individuals susceptible to disease Y, identified in the screening procedure, a treatment is available that results in no adverse health effects and costs \$2,000. The screening process itself costs \$30 per person.

Suppose there are 100,000 individuals who might be screened annually. Without screening, 1 percent, or 1,000 individuals, develop disease Y during the year, costing \$10,000,000 and 50 QALYs (measured relative to a world in which there was no disease Y). The screening may catch 30 percent of those who would develop disease Y (with no Type II errors—i.e., no false positives). This means that 700 individuals still get disease Y, costing \$7,000,000 and 35 lost QALYs, while 300 are treated early at a cost of \$600,000 and no QALY impacts. Screening 100,000 individuals costs \$3,000,000. So, in total, the screening intervention has a cost of \$10,600,000 and 35 lost QALYs (again, compared to a world in which there was no disease Y). Compared with the status quo of no screening, screening results in a net cost of \$600,000 while saving 15 QALYs. The cost-effectiveness ratio of screening is \$40,000 per QALY, a figure that most would consider an acceptable, although not outstanding, health return for the intervention.

Now, though, suppose that the effectiveness of the screening procedure turns out to be optimistic. Suppose that instead of identifying 30 percent of those who would develop disease Y, it identifies only 25 percent. Further, suppose that for 0.5 percent of those who would not develop disease Y, the test provides a false positive, leading to the \$2,000 treatment regime being used on these individuals, too, even though it is not needed.

In this case, out of the pool of 100,000 individuals, 750 still develop disease Y, costing \$7,500,000 and 37.5 QALYs; 250 who would develop disease Y are caught early and treated, at a cost of \$500,000; and 500 who would not have developed disease Y are treated anyway, at a cost of \$1,000,000. Now, compared to the status quo, the net cost of screening is \$2,000,000 while the net savings in health is 12.5 QALYs. The cost-effectiveness ratio is \$160,000 per QALY.

In this case, then, two modest sources of error lead to a four-fold increase in the cost per QALY. An original analysis would be quite misleading if it did not indicate that the conclusions about a CUA would be quite strongly affected by small changes in the effectiveness of the screening process. While this example may seem somewhat rigged, its point is not. For many types of health interventions, the degrees of uncertainty present may qualitatively affect the outcome of an analysis.

## **2.3. Sensitivity Analysis**

The most common way of examining the impacts of uncertainty on the conclusions of a CUA is sensitivity analysis. As in the example just described, sensitivity involves doing separate analyses for different values of a variable that may be uncertain and then examining the effects of the different analyses on the overall conclusions.

Sensitivity analysis is often very useful. It can indicate which variables are such that small changes in their assumed values can have a large impact on the conclusions of an analysis and which variables can change substantially without affecting the conclusions in any significant way. Variables for which small changes can lead to large changes in the conclusion can then be singled out. A range of possible values for these variables can be adopted and the resulting range of possible outcomes reported. Sensitivity analysis also allows a researcher to direct attention to the need to gather more information about certain variables in order to reduce their uncertainty.

In principle, values for multiple variables can be modified simultaneously, although this is seldom done in practice. One could, however, easily imagine a table indicating, along the horizontal and vertical margins, different values for two important variables in the analysis and reporting the resulting cost-effectiveness ratio for each pair of values. This sort of approach allows sensitivity analysis to take account of multiple uncertainties, at least in a limited way.

#### **2.4. Break-even Analysis**

Similar to sensitivity analysis is break-even analysis. It is particularly useful when there is one key uncertain variable in an analysis. Break-even analysis involves "working backward" to determine the cutoff value for the key uncertain variable which would make the overall intervention cost effective or not.

For instance, consider a screening program for cholesterol. Suppose there is some knowledge of the costs of heart disease; of the impact of heart disease on mortality, morbidity, and the quality of life; of the costs of the screening program; and, even concerning the distribution of cholesterol readings one is likely to encounter in a given population to be tested. However, the link between an extra 10 cholesterol points and the increased probability of heart disease may be very difficult to determine. In such cases, one might well work backward and derive the threshold level for this probability in order to lead to a cost-effectiveness ratio below perhaps \$40,000 per QALY. This break-even level may be such that experts are willing to say, "almost certainly the link is that strong," or "it is unlikely the link is that strong." In either case, some solid sense can be reached about the cost effectiveness of the intervention without having to determine with precision the value for the uncertain variable.

#### **2.5. Best-Case/Worst-Case Analysis**

Another approach for explicitly handling uncertainty in CUA is best-case/worst-case (BCWC) analysis. In this approach a set of uncertain variables are assigned "optimistic" and "pessimistic" values. Optimistic values result in a low cost-effectiveness ratio, pessimistic values in a high ratio. A CUA is done with both sets of values.

This approach is often very useful, especially for situations in which there are multiple sources of uncertainty. However, it also has some drawbacks. There is some lack of precision in the notion of a "worst-case" or "best-case" value. Usually, the intent is not to pick literally the absolute worst or best value that might conceivably occur but rather some "reasonably optimistic" and "reasonably pessimistic" assumptions; again, the term "reasonably" is somewhat vague.

Additionally, if there are many stages in the analysis, the permutation of pessimistic or optimistic assumptions may give unlikely results. For instance, suppose a cost-effectiveness ratio involves the product of three variables, X, Y, and Z. Suppose there is uncertainty about the values for each of these variables and, therefore, uncertainty about the value of their product, XYZ. Let a "reasonably pessimistic" value for a variable mean a value that is at the 95-percent level of that variable's

cumulative distribution function—i.e., only 5 percent of the variable is above this reasonably pessimistic value. If each variable is assigned a reasonably pessimistic value in this way, then, assuming the variables are uncorrelated, the resulting reasonably pessimistic value for XYZ will be at approximately the .9999 level of the cumulative distribution function for XYZ. This means that there is only about a .01 percent chance that XYZ is above this value.

Clearly, as in this example, the combination of "reasonably pessimistic" values for several variables at once can lead to a conclusion that has moved beyond reasonably pessimistic and to extraordinarily pessimistic. This type of problem can make the resulting BCWC analysis somewhat difficult to interpret. On the other hand, if a BCWC analysis indicates that the overall conclusions are not very sensitive to different values for uncertain variables, then added confidence is given about the robustness of the overall "base-case" conclusion of the CUA.

## **2.6. Formally Keeping Track of the Uncertainties in a CUA**

One disadvantage of the above approaches to handling uncertainty in CUA is that an explicit, numerical indication of the overall uncertainty associated with a cost-effectiveness ratio is not presented. For many purposes, it would be highly desirable to report something akin to a standard error for the overall cost-effectiveness ratio. This can be done if care is taken in explicitly representing the uncertainty that characterizes each variable in the analysis. In essence, in this approach, information about the uncertainty for each variable is used to compute the net uncertainty associated with the final conclusion of the analysis.

One problem in pursuing this approach is that it appears difficult analytically. It is not easy to combine variables in complex ways and, considering each variable to be characterized by a given probability distribution, write down the probability distribution for the combined product. This is true even for simple and well-behaved probability distributions like a normal or Poisson or uniform distribution, let alone for the more complex probability distributions.

Fortunately, there is a short-cut of sorts that can be employed to combine probability distributions. It is known as Monte Carlo analysis. To understand Monte Carlo analysis, consider a computer package that can generate sample values for a variable that has a given probability distribution. Such random number generators are available in almost all statistical software packages. To illustrate how the random number generator works, suppose a variable  $X$  is distributed normally with mean 10 and standard deviation of 2.5. Then, the random number generator would provide for this variable a sample value which has these features. That is, if a large number of such values for  $X$  were generated by this process, the resulting sample would have a mean close to 10, a standard error close to 2.5, and a normal distribution. It can be shown that the larger the size of the sample, the closer the sample mean and standard deviation theoretical values will be to their theoretical counterparts.

Now suppose that for all the uncertain variables in a CUA, a given value for each is generated in a fashion analogous to that described above. These values are then used to compute an associated overall cost-effectiveness ratio. This cost-effectiveness ratio is not the true cost-effectiveness ratio. It is the value for the cost-effectiveness ratio that would prevail if each of the uncertain variables happened to assume a value equal to that generated by the computer.

However, if this process is repeated over and over, each time producing a sample cost-effectiveness ratio, the resulting range of cost-effectiveness ratios will begin to look like the probability distribution for the true cost-effectiveness ratio. In general this process will lead to a sample distribution for the overall cost-effectiveness ratio that, as the sample size increases, converges to the true distribution.

This process may seem very difficult. If all the analysis were to be done by hand, it surely would be. However, it is simple to instruct even the most basic personal computer to do this type of analysis quickly.

As an example of this approach, consider a cost-effectiveness ratio of  $X/Y$  where  $X$  is costs and  $Y$  is quality-adjusted life years. Let QALY be a function,  $f(\cdot)$ , of a clinical measure,  $Z$ . So,  $f(\cdot) = f(Z)$ . Suppose there is uncertainty with regard to  $X$ , with regard to  $Z$ , and with regard to  $f(\cdot)$ .

Suppose that costs,  $X$ , are distributed normally with costs of \$100,000 and a standard deviation of \$10,000. Let  $Z$  have a more complicated distribution. Suppose it is judged that there is a 25 percent probability that the health intervention has no impact at all. If it does have a positive benefit, the level of benefit is uncertain. Suppose this benefit, conditional on some positive benefit, is judged to be exponentially distributed. An exponential distribution has a density function that begins at 0, rises to a peak, and then declines with a long tail to the right. There is one parameter in an exponential distribution, typically known as "beta." Here let beta be .225, which implies that the mean of the exponential distribution is 4.44 (and its standard deviation is 4.44 as well, since the mean and standard deviation of an exponential distribution are the same). Finally, let there be a uniform distribution between .95 and 1.05 of the link between  $Z$  and  $Y$ . So  $Y$  equals  $W \times Z$  where  $W$  is drawn from a uniform distribution between .95 and 1.05.

The outcome of interest is  $X/Z$ . Given the independence we have assumed between  $X$  and  $Z$ , it is not difficult to derive a point estimate for  $X/Y$ , which is \$30,000 per QALY. To see this, note that there is a .25 probability that  $Z$  is 0 and a probability of .75 that  $Z$  is positive. The expected value of  $Z$  if it is positive is 4.44. So, the expected value of  $Z$  is  $(.25 \times 0) + (.75 \times 4.44)$ , or 3.33. The expected value of  $W$  is 1, and since  $Y = W \times Z$ , the expected value of  $Y$  is therefore also 3.33. Finally, the expected value of  $X$  is \$100,000. So, the expected value of  $X/Z$  is  $\$100,000/3.33$  or \$30,000. But accurately gauging the uncertainty in this estimate may seem difficult. Deriving it analytically would be so, but it is not difficult with a computer. The result of a short computer program using the SAS statistical package indicates that the cost-effectiveness ratio is distributed with a mean of \$30,008 with a standard deviation of \$2,524. This program involved only six lines of instructions and took less than a second to run on a personal computer.

It can be noted that best-case/worst-case analysis would not be very appropriate for this example. Indeed, in the worst-case analysis there is no health benefit and the cost-effectiveness ratio would be infinity. This would suggest much more uncertainty than is actually present. It can also be noted that the example assumes that each of the uncertainties were independent. It would be easy to incorporate views about dependence into the analysis and, in general, one should consider such dependencies for any CUA.

The main problem in this approach is that for each key variable, the precision or confidence of the conclusions must be weighed and an explicit probability distribution provided. On occasion, there may be objective information about the uncertainty characterizing some variable, and this objective information can be used to provide an explicit probability distribution for the variable. This might result, for instance, if information is used in a CUA that comes from a controlled experiment examining the clinical effectiveness of a treatment. The statistical analysis from the controlled experiment will generally result in not only an estimate of the clinical effectiveness of the treatment but also an estimate of the uncertainty characterizing the effectiveness estimate. An example is a standard error associated with a coefficient estimate in a regression analysis.

In other cases, however, the primary source of information about the level of a variable and the uncertainty characterizing its level is expert opinion. Expert opinion does not magically make up for the lack of good objective research findings, but this does not mean we should ignore it.

A growing field in applied decision analysis specializes in eliciting subjective probability estimates from experts. Specialized techniques have been developed to try to ensure that such expert probability judgments are well calibrated—i.e., that they accurately translate experts' beliefs into probability values—and that they avoid the common errors identified in the psychology literature.

Applied decision analysis has also developed techniques for overcoming a natural reluctance of scientists unfamiliar with this approach to express their subjective views in quantitative terms. For instance, in eliciting prior beliefs about the impact of acid rain on the health of fish, an expert may initially say, "we do not have enough information to make any statements about that." After probing, however, it emerges that there is more to be said than this. The expert is willing to say that some things are known about the impact of acid rain on fish. He or she may be willing to say that it is likely that the effect is not as large as some given value, as small as some other value, and so forth. Before long, a subjective probability estimate emerges.

If the resulting probability distribution over the impact of a health intervention is so large that the CUA displays substantial uncertainty, then so be it. The purpose, after all, is not to get rid of uncertainty in a CUA but to indicate accurately its impact on the conclusions of a CUA.

Different experts may have dissimilar views about the likely impact of a given health intervention. These differences of opinion again do not invalidate CUA or the approach of gathering subjective probability distributions. If the conclusion of a CUA is completely dependent on whose beliefs about key variables are believed, then again, so be it. The proper conclusion to the CUA, in that case, is that the conclusion is totally dependent on which expert is believed.

## **2.7. Concluding Comments**

Despite these methods for dealing with uncertainty in CUA, the basic point remains: there are cases in which the degrees of uncertainty are so extreme that there is little that can be said in terms of cost effectiveness of a health intervention. CUA should be designed to help indicate when this is so. It should be a technique that, as a matter of course, accounts for uncertainties to the degree possible and is open to finding uncertainties even if they make it difficult to draw substantive conclusions about the cost effectiveness of a health intervention.

Another point deserves reiteration: It is important not to fall victim to the aversion of the political and bureaucratic system for dealing with uncertainty. Policymakers want a single, concrete bottom line. Sometimes, however, the answer has got to be "it is not that simple," or "it depends." Practitioners of CUA should not be afraid to deliver that message when it is warranted.

## **3. DISCOUNTING**

### **3.1. Introduction**

In CUA, the costs and health effects from a health intervention often occur over an extended period of time. There is substantial controversy about how to account for costs and health effects that occur in the future. The issues at debate are often quite complex and will not be developed in extensive

detail here. Instead, an overview will be given of common practice and some key areas of contention.

### 3.2. The Mechanics of Discounting

The most frequent way of handling future costs and health effects is through discounting. In almost all CUAs, costs are discounted. In most, health outcomes (or quality-of-life outcomes) are also discounted, usually at the same discount rate. The mechanics of discounting are relatively straightforward, especially if there is a given discount rate that remains fixed over time.

Let  $(C_0, C_1, \dots, C_T)$  represent a *time stream* of costs; during the present period the costs are  $C_0$ , during the next period the costs are  $C_1$ , and so forth, until in  $T$  periods from now the costs are  $C_T$ . Let  $r$  be the per period *discount rate*. Then the *net present value* (NPV) of this time stream of costs is:

$$(11-1) \text{NPV}(C_0, C_1, \dots, C_T) = \sum [C_t / (1+r)^t].$$

So, for instance, if  $r = .03$ , then the NPV of (10,5,8,11,4,6) is:

$$\begin{aligned} 10/(1.03)^0 + 5/(1.03)^1 + 8/(1.03)^2 + 11/(1.03)^3 + 4/(1.03)^4 + 6/(1.03)^5 &= \\ 10/1 + 5/(1.03) + 8/(1.0609) + 11/(1.0927) + 4/(1.1255) + 6/(1.1593) &= \\ 10 + 4.854 + 7.541 + 10.067 + 3.554 + 5.176 &= 41.192, \end{aligned}$$

where the figures have been rounded to four decimal points in the second line and three decimal points in the third line. This NPV figure is less than the figure—44—that would be obtained by simply adding the different costs together. Discounting results in lower figures for future effects.

In the above example, all of the  $C_t$  are negative. This is not necessary for the above formula to be valid. One can similarly discount a time stream in which some entries are positive and some negative.

In considering a time stream of costs, a natural issue is the length of "one unit of time." For instance, suppose that in the above example each unit of time is 1 year and, therefore, the discount rate of .03 is an annual discount rate. What would happen if, instead of measuring costs in 1-year intervals, we chose to measure costs in half-year intervals. In this case, the time stream (10,5,8,11,4,6) might become (5,5,2.5,2.5,4,4,5.5,5.5,2,2,3,3).

Suppose the discount rate is divided in half as well, now being .015. Would reapplying Equation (11-1) result in the same calculation? The results would be close but not exactly the same. In the case where all the  $C_t$  are positive, the result would be a NPV value that is slightly smaller. Indeed, dividing the time period (and the discount factor) more and more finely, the NPV value continues to change, the changes becoming smaller and smaller. It can be shown that in the limit, as the time period and discount rate are divided more and more finely, the following formula holds:

$$(11-2) \text{NPV} = \int e^{-rC(t)} dt,$$

where  $C(t)$  is a continuous variable measuring the cost occurring at time  $t$ . This is the typical formula for what is called *continuous discounting*.

If data are available that measure costs very finely as a function of time, Equation (11-2) is the proper way to compute the present discounted value of the time stream of costs. If data are available only in

more "lumpy" time intervals, the formula (11-1), using a year as the unit of time, is often used as an alternative and in general provides a quite acceptable approximation to (11-2).

In moving from discounting of costs alone to discounting both costs and health outcomes, both the numerator—costs—and the denominator—QALYs—can be represented as time streams and discounted in a manner analogous to that considered above. Expressed in discrete time units, let the cost time stream be denoted as  $(C_0, C_1, \dots, C_T)$ , and let the quality-of-life time stream be denoted as  $(QOL_0, QOL_1, \dots, QOL_T)$ . Let  $r_c$  be the discount factor for costs and let  $r_q$  be the discount factor for quality-of-life outcomes. Then, the cost-effectiveness ratio is computed with the presented discounted value of costs in the numerator and the presented discounted value of quality-of-life impacts in the denominator. Or, the cost-effectiveness ratio is:

$$(11-3) \quad \{\Sigma [C_t/(1+r_c)^t]\} / \{\Sigma [QOL_t/(1+r_q)^t]\}.$$

Measured in continuous time, the cost-effectiveness ratio becomes:

$$(11-4) \quad \int e^{-r_c C(t)} dt / \int e^{-r_q QOL(t)} dt.$$

### 3.3. The Rationale for Discounting Costs

If an individual receives \$1 today, he can put it in the bank and receive interest. Given the choice between receiving \$1 today or \$1 in  $N$  years from now, the individual should clearly prefer receiving \$1 today. He or she can put it in the bank and, as indicated, have more than \$1 in  $N$  years. Put differently, money received in the future is worth less than the same amount of money received in the present.

Suppose the individual not only can save money and receive a rate of return but also can borrow money from the bank. Suppose also that he or she can borrow money at a rate of interest of  $r$ , the same interest rate he receives for saving money. Then, by analogous reasoning, the individual will prefer paying \$1 in the future to paying \$1 today. Put differently, money spent in the future is not as "costly" to the individual as the same amount of money spent today.

Discounting, using the formulas from the prior subsection, is the procedure that the rational individual facing these interest rates should apply to determine how monetary costs and benefits in the present and future should be taken into account and compared with one another. He will evaluate any time stream of monetary costs and benefits based on its present discounted value, using these formulas.

In moving from an individual to a societal perspective, the justification for discounting is similar. Assuming that saved resources can be put to productive use, a dollar's worth of resources saved today can generate more than a dollar's worth of resources tomorrow. As such, the same sort of discounting procedure applies for aggregate costs as for individual costs.

While this chain of reasoning is widely accepted in CUA and other forms of economic analysis, even here some complications arise. For instance, in applying discounting at an aggregate, societal level, different individuals will be alive to bear the costs depending on when they are incurred. Distributional implications across generations are ignored in discounting. Suppose one generation borrows \$1,000,000,000. If the discount rate is 3 percent, then in 20 years a debt of \$1,806,000,000 will exist. Is it really immaterial whether one generation borrows the money and leaves it to the next to pay the bill? Discounting would indicate that either paying \$1,000,000,000 today to cover the debt or paying \$1,806,000,000 in 20 years is equally desirable. If the debt is deferred, the

\$1,000,000,000 that the first generation no longer has to pay today can be put in the bank, passed down to the next generation, and in 20 years be worth the \$1,806,000,000 required to pay off the bill. But what if the money is used to benefit the earlier generation and no \$1,000,000,000 is actually put in any bank account for the future generation? Such considerations are not taken into account in discounting.

A more practical problem is inflation, although it is easily handled. Before discussing this, two definitions are in order. "Nominal" costs refer to costs measured in terms of the number of dollars involved at the time the cost is incurred. "Real" costs translate this nominal figure into the amount of dollars that the same item would have cost in some base year. So, for example, an apple might cost \$.25 in 1970 but, because of inflation, costs \$.50 in 1991. The nominal cost of the apple in 1991 is \$.50. The real cost of the apple in 1991, measured in 1970 dollars, is \$.25. The most common method for dealing with the effects of inflation is to measure both costs and discount rates in real terms. It is acceptable to use both nominal costs and a nominal discount rate; the same results will be obtained. It is not acceptable, however, to use nominal costs and a real discount rate or a nominal discount rate and real costs.

While acceptable to use nominal costs and a nominal discount rate, the recommended procedure is to use real discount rates. This is because it is very difficult to predict inflation and therefore to predict future nominal costs and future nominal interest rates. In contrast, it is much easier to predict real costs and real interest rates, because these are much more stable over time than their nominal counterparts.

Another practical complication is that taxation puts a wedge between the real, after-tax interest rates that individuals pay when they borrow and receive when they save funds. For instance, suppose an individual can borrow or save at a 10 percent interest rate before taxes. If he saves \$100, after a year he will have \$110. But, the \$10 in interest will be taxed at, say, 28 percent, meaning that his after-tax rate of return is only 7.2 percent. In contrast, if he cannot deduct his interest on borrowed funds, his after-tax interest rate for borrowing funds remains at 10 percent.

This is an example of how the assumption of "perfect capital markets," assumed in the simple version of discounting, is violated in practice. Other violations of the perfect capital market assumption can also be provided. For example, in reality different people face different interest rates because their assets and credit ratings differ, because they borrow and save at different rates, and because interest rates change somewhat over time. The interest rate one faces, in fact, is actually a whole spectrum of interest rates depending on the length of time one's money must be tied up and on the risk of the savings vehicle.

All of these concerns make it difficult to determine the "one correct" interest rate that should be used in discounting and, indeed, there is not one common rate that is accepted across studies. Different studies use different rates. It is therefore recommended that several rates be used to determine the sensitivity of the results to different discounting assumptions.

If real interest rates for safe, short-term investments are used to determine the discount rate, then a discount rate of 3 percent or so would seem appealing. This is close to the historical rate of return on 1-year U.S. Treasury Bills. On the other hand, real interest rates in the United States have generally been high in the past decade, and it could be argued that a rate closer to 4 or 5 percent would be more appropriate.

From a slightly different perspective, some recommend using the rate of productivity growth of the economy for the discount rate in economic studies, since this indicates the extent to which current savings can be converted into future real value by the economy as a whole. This would lead to a discount rate closer to 2 percent.

Some subscribe to using productivity levels to gauge the appropriate discount rate but argue that, since CUA in health care analyzes health interventions, productivity levels in health are the appropriate guide. This leads to additional complications. For instance, if one anticipates tremendous productivity advances in medical care, then a dollar saved today allows for the purchase of medical care resources in the future that accomplish much more in terms of health benefits than if spent today. This would imply a much higher discount rate.

Taking all this into account, it is recommended that computations of cost-effectiveness ratios use real discount rates of 0 percent (as a base-case for purposes of comparison), 3 percent, and 5 percent. This recommendation is admittedly somewhat arbitrary but seems to reflect reasonably well the range of practice commonly reported in the current literature.

### **3.4. The Rationale for Discounting QALYs**

If there is controversy about discounting costs, it pales beside controversies about discounting QALYs. Weinstein and Stason (1976) strongly advocate discounting future health effects using the same discount rate used to discount costs. Others object to any form of discounting of health benefits, feeling it discriminates against the young and against future generations.

The issues involved in the debate are complex. They involve questions about individuals' intertemporal time preferences, individual preferences about the duration of being in a given health state (something that may vary across health states and between morbidity and mortality considerations), and intergeneration or cross-cohort distributional concerns. As Lipscomb (1989) argues, any single discount rate for health effects merges these concerns, often in some ambiguous fashion.

As indicated above, we do not dwell on the theoretical issues of debate at length here. Instead, a set of analyses is suggested that accommodates readers of different persuasions on this issue. Thus, it is recommended that a CUA report cost-effectiveness ratios for the following combinations of real discount rates (using real cost figures):

- (1) Discount costs at 0 percent and health effects at 0 percent.
- (2) Discount costs at 3 percent and health effects at 0 percent.
- (3) Discount costs at 5 percent and health effects at 0 percent.
- (4) Discount costs at 3 percent and health effects at 3 percent.
- (5) Discount costs at 5 percent and health effects at 5 percent.

If the preferred single procedure among these is to be judged on the basis of frequency of use in empirical analysis, approach (4), discounting costs and health effects at a real rate of 3 percent, would probably be close to the modal approach in the literature.

Because some believe that discounting health effects discriminates against the young and future generations, it is recommended that additional analyses be provided, whenever costs and health outcomes are borne by substantially different age cohorts. In that case, the cost-effectiveness ratio should probably be computed separately for different age cohorts and reported along with the aggregate cost-effectiveness ratio. It is not clear that CUA at its current state of the art can resolve distributional issues of allocating costs and health benefits across generations, but when these issues are of particular relevance for a health intervention, CUA can at least make clear the implications for each age cohort.



## CONCLUSION

This monograph has provided an introduction to cost-utility analysis as an evaluative tool for health care interventions. CUA is by no means the only useful economic approach for evaluating health care interventions, but it has increasingly become one of the most important and most valuable. CUA continues to evolve toward a commonly-accepted set of steps, procedures, and interpretation. It has come a long way in a relatively brief period in this regard. At the same time, there remain controversies and unresolved issues in CUA. It is hoped that this monograph has not failed to give the reader some sense of these as well.

It is important to recognize that CUA cannot make up for a lack of understanding about the basic nature of a given health care intervention. It cannot make a judgment about the cost effectiveness of a health care intervention when the existing clinical knowledge about the intervention does not provide sufficient information about the intervention's effectiveness, its impacts on the patient's quality of life, or the direct and other costs incurred. This does not mean that some degree of uncertainty about these factors cannot be handled within the context of CUA. Nor does it mean that CUA is not useful even in cases when uncertainty about these key variables is extensive. Even in these cases, CUA at least provides a framework for identifying the types of data about which additional information is required. Still, CUA cannot substitute for a valid scientific understanding of the basic nature of a health care intervention. In many instances, this, rather than any controversy in how to conduct a CUA, is the weak link in being able to evaluate a health care intervention from an economic and social perspective.

Despite these caveats and potential limitations, CUA presents a logical, powerful, and objective tool. In trying to promote given health interventions or to accentuate the seriousness of a given illness, there are often incentives to make the monetary numbers—which individuals such as legislators, public executives, and news media readily understand—be as large as possible. This is particularly the case in cost-of-illness analyses in which the "game" is often to make the illness under consideration appear to the public or to politicians and policymakers to have tremendous costs to society. This, in turn, is used to justify and promote increased levels of funding for research, prevention, and treatment programs. On occasion, this process is turned on its head, and interested parties have an incentive to minimize the seriousness of a given illness or the value of a given health care intervention.

In the battle for scarce resources, this process is ultimately self-defeating as programs use the form of analysis and types of figures that serve their own parochial ends and, if anything, policymakers become increasingly skeptical of them all. It has been the perspective of this work that what is needed is not a tool of evaluation that is designed to promote given treatment or prevention programs, but to provide legitimate and compelling guidelines which, if adopted, will provide a "level playing field" for all potential health interventions. CUA, though still in its youth, has the potential to provide decision makers with this much needed perspective from which to view competing claims for health care resources.



## REFERENCES

- Anderson JP, Bush JW, Chen M, Dolenc D, Policy Space Areas and Properties of Benefit-Cost/Utility Analysis, *Journal of the American Medical Association*:795, 1986.
- Andreason NC, Winokur G, Secondary Depression: Familial, Clinical or Research Perspectives, *American Journal of Psychiatry* 136:62-6, 1979.
- Bartel A, Taubman P, Health and Labor Market Success: The Role of Various Diseases, *Review of Economics and Statistics* 20:1-8, 1979.
- Bartel A, Taubman P, Some Economic and Demographic Consequences of Mental Illness, *Journal of Labor Economics* 4:243-56, 1986.
- Benham L, Benham A, Employment, Earnings, and Psychiatric Diagnosis, in V. Fuchs (ed) *Economic Aspects of Health*, Chicago:University of Chicago Press, 1981, pp. 203-20.
- Bothwell S, Weissman MM, Social Impairments Four Years After an Acute Depressive Episode, *American Journal of Orthopsychology* 47:231-7, 1977.
- Boyd JH, Weissman MM, Epidemiology of Affective Disorders. *Archives of General Psychiatry*, 38:1039-46, 1981.
- Boyd NF, Sutherland HJ, Ciampi A, Tibshirani R, Till JE, Harwood A, A Comparison of Methods of Assessing Voice Quality in Laryngeal Cancer, in: R.B. Deber and G.G. Thompson, eds., *Choices in Health Care: Decision Making and Evaluation of Effectiveness*, Toronto:University of Toronto, Department of Health Administration, 1982.
- Buchanan JM, Tullock G, *The Calculus of Consent*, Ann Arbor: The University of Michigan Press, 1962.
- Buchanan J, Vanberg V, A Theory of Leadership and Deference in Constitutional Construction, *Journal of Constitutional Economics*, forthcoming, 1990.
- Bush JW, Fanshel S, Chen MM, Analysis of a Tuberculin Testing Program Using a Health Status Index, *Socioeconomic Planning Sciences*, 6(1):49-69, 1972.
- Campbell A, Converse PE, Rodgers WL, *The Quality of American Life*, New York: Russell Sage Foundation, 1976.
- Churchill N, Morgan H, Torrance GW, Quality of Life in End-stage Renal Disease, *Peritoneal Dialysis Bulletin* 4:20-3, 1984.
- Clayton PJ, Halikas JA, Maurice WL, The Depression of Widowhood, *British Journal of Psychiatry* 121:71-8, 1972.
- Daniels N, *Just Health Care*, Cambridge, MA:Cambridge University Press, 1985.
- Deniston L, Carpentier-Alting P, Kneisley J, Hawthorne M, and Port FK, Assessment of Quality of Life in End-Stage Renal Disease, *Journal of Health Services Research*, 24(4):555-78, October 1989.

- DiMascio A, Weissman MM, Prusoff BA, Neu C, Zwilling M, Klerman GL, Differential Symptom Reduction by Drugs and Psychotherapy in Acute Depression. *Archives of General Psychiatry*, 36:1450-6, 1979.
- Doubilet P, Weinstein MC, McNeil BJ, Use and Misuse of the Term "Cost Effective" in Medicine, *New England Journal of Medicine*, 314(4), 253. 253(15):2229-35, 1985.
- Drummond MF, Stoddart GL, Torrance, GW, *Methods for the Economic Evaluation of Health Care Programmes*, New York:Oxford University Press, 1987.
- Emery DC, Schneiderman LJ, Cost-effectiveness Analysis in Health Care, *Hastings Center Report*, July/August 1989.
- Fischhoff B, Firby L, Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility, *Journal of Risk and Uncertainty*, forthcoming.
- Fischer GW, Utility Models for Multiple Objective Decisions: Do They Accurately Present Human Preferences? *Decision Sciences* 10, 1979.
- Frank RG, Gertler PJ, An Assessment of Measurement Error Bias for Estimating the Effect of Mental Distress on Income, manuscript, 1988.
- Frank RG, Kamlet MS, Provider Choice for the Treatment of Mental Disorder: The Impact of Health and Mental Health Status, *Journal of Health Services Research*, 1989.
- Fryback DG, Thronbury JR, The Efficacy of Diagnostic Imaging, *Medical Decision Making*, Vol 11, No. 2, 1991.
- Gibbard A, The Prospective Pareto Principle and Equity of Access to Health Care, in President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, *Securing Access to Health Care: The Ethical Implications of Differences in the Availability of Health Services*, Volume Two: Appendices, Appendix G, Washington D.C.: U.S. Government Printing Office, 1983.
- Gramlich E, *Benefit-Cost Analysis of Government Programs*, Englewood Cliffs, NJ: Prentice-Hall, 1981.
- Hall JA, Epstein AM, McNeil BJ, Multidimensionality of Health Status in an Elderly Population: Construct Validity of a Measurement Battery, *Medical Care*, 27(3) Supplement:S168-S76, March 1989.
- Hall JA, Epstein AM, McNeil BJ, Multidimensionality of Health Status in an Elderly Population: Construct Validity of a Measurement Battery, *Medical Care*, 27(3) Supplement:S168-S76, March 1989.
- Hatziandreu EI, Koplan JP, Weinstein MC, Caspersen CJ, Warner K, A Cost-effectiveness Analysis of Exercise as a Health Promotion Activity, *American Journal of Public Health*, 78(11), 1988.
- Jacobs P, *The Economics of Health and Medical Care*, Second Edition, Rockville, MD:Aspen, 1987.
- Johnson EJ, Tversky A, Affect, Generalization, and the Perception of Risk, *Journal of Personality and Social Psychology*, 45(1), 20-31, 1983.

Jones-Lee MW, *The Value of Life*, Chicago: University of Chicago Press, 1976.

Kahneman D, Tversky A, Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, 47, 263-91, 1979.

Kaplan RM, Bush JW, Health Related Quality of Life Measurement for Evaluation Research and Policy Analysis, *Health Psychology* 1(1): 61-80, 1982.

Kaplan RM, Bush JW, Berry CC, Health Status: Types of Validity and the Index of Well-being, *Journal of Health Services Research*, 11(4): 478-507, 1976.

Kaplan RM, Anderson JP, A General Health Policy Model: Update and Applications, *Journal of Health Services Research*, 23(2):203-35, 1988.

Keeney R, Raiffa H, *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*, New York: Wiley, 1976.

Keller MB, Shapiro RW, Lavori PW, Wolfe N, Recovery in Major Depressive Disorder: Analysis with Life Tables and Regression Models, *Archives of General Psychiatry* 39:905-10, 1982a.

Keller MB, Shapiro RW, Lavori PW, Wolfe N, Relapse in Major Depressive Disorder: Analysis with Life Table, *Archives of General Psychiatry*, 39:905-10, 1982b.

Klerman GL, Dimascio A, Weissman M, Prusoff B, Paykel ES, Treatment of Depression by Drugs and Psychotherapy, *American Journal of Psychiatry* 131:186-91, 1974.

Klerman GL, Long-Term Treatment of Affective Disorders, In Lipton, M.A., DiMascio, A. and Killam, K.F. (eds), *Psychopharmacology: A Generation of Progress*, New York:Raven Press, 1978, pp. 1303-1311.

Klerman GL, Weissman MM, Roundsaville BJ, Chevron ES eds., *Interpersonal Psychotherapy of Depression*, New York:Basic Books, 1984.

Klarman HE, Present Status of Cost-Benefit Analysis in the Health Field, *American Journal of Public Health*, 57, November 1967, 1948-53. *Preventive Medicine* 7:414-423, 1978.

Lave JR, Lester B, Cost-benefit Concepts in Health: Examination of Some Prevention Efforts, *Preventive Medicine* 7:414-23, 1978.

Lipscomb J, Time Preferences for Health in Cost-Effectiveness Analysis, *Medical Care*, 27(3), S233-S53, 1989.

Lloyd C, Life Events and Depressive Disorder Reviewed," *Archives of General Psychiatry*, 37:541-8, 1980.

McNeil BJ, Weichselbaum R, Pauker S, Tradeoffs between Quality and Quantity of Life in Laryngeal Cancer, *New England Journal of Medicine* 305(17):982-7, 1981.

- Miller T, Costs of Increasing the Speed Limit to 65 MPH, The Urban Institute Policy and Research Reports 19(2):22 Summer 1989.
- Mishan EJ, *Cost-Benefit Analysis: An Informal Introduction*, Fourth Edition, Unwin Hyman, London, 1988.
- Mishan EJ, *Evaluation of Life and Limb*, *Journal of Political Economy*, 79, 687-705, July-August 1971.
- Monkoff K, Bergman E, Ceck A, et al. Hopelessness, Depression, and Attempted Suicide, *American Journal of Psychiatry*, 130:455-9, 1973.
- Morgan RF ed, *The Iatrogenics Handbook: A Critical Look at Research and Practice in the Helping Professions*, Toronto, Ontario:IPI Publishing, 1983.
- Mushkin SJ, Toward a Definition of Health Economics, *Public Health Reports*, 79, 785-93, July/August 1958.
- Myers JK, Weismann MM, Trischler GL, et al. Six-month Prevalence of Psychiatric Disorders in Three Communities: 1980-1982, *Archives of General Psychiatry*, 41:959-67, 1984.
- Office of Technology Assessment, U.S. Congress, The Implications of Cost-effectiveness Analysis of Medical Technology, Washington, DC:U.S. Government Printing Office, September 1980.
- Office of Technology Assessment, U.S. Congress, *A Review of Selected Federal Vaccine and Immunization Policies: Bases on Case Studies of Pneumococcal Vaccine*. Washington, D.C.: U.S. Government Printing Office, 1979.
- Patrick DL, Bush JW, Chen MM, Methods for Measuring Levels of Well-being for a Health-status Index, *Journal of Health Services Research* 8(3):228-45, 1973.
- Pliskin JS, Shepard DS, Weinstein MC, Utility Functions for Life Years and Health Status, *Operations Research* 28(1):206-24, 1980.
- Prusoff BA, Weissman MM, Klerman GL, Roundsaville BJ, Research Diagnostic Criteria Subtypes of Depression. Their Role as Predictors of Differential Response to Psychotherapy and Drug Treatment. *Archives of General Psychiatry* 37:796-801, 1980.
- Rapopart J, Robertson R, Stuart B, *Understanding Health Economics*, Rockville, MD:Aspen, 1982.
- Rice DP, *Estimating the Cost of Illness*, Washington, D.C.: U.S.Government Printing Office, 1966.
- Rogers PJ, Eaton EK, Bruhn JG, Is Health Promotion Cost-effective? *Preventive Medicine* 10:324-39, 1981.
- Rosenzweig MR, Schultz PT, Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight, *Journal of Political Economy*, 91(5), October 1983.
- Russell LB, *Is Prevention Better than Cure*, Washington, D.C.:Brookings Institution, 1986.

- Russell LB, *Evaluating Preventive Care: Report on a Workshop*, Washington, D.C.: The Brookings Institution, 1987.
- Russell LB, Taylor W, Fahs M, Rose D, Komaroff A, Mandelblatt J, Pass T, Schechter C, Shepard D, Silver A, *Some Examples of the Cost-effectiveness of Preventive Services: Background Paper for the U.S. Preventive Services Task Force*, unpublished, 1989.
- Sackett DL, Torrance GW, The Utility of Different Health States as Perceived by the General Public, *Journal of Chronic Diseases* 31:697-704, 1978.
- Schelling TC, The Life You Save May Be Your Own, in S.B. Chase, Jr., ed., *Problems in Public Expenditure Analysis*, Washington D.C.: Brookings Institution, 1968, 127-162.
- Shepard DS, Thompson MS, First Principles of Cost-effectiveness Analysis in Health, *Public Health Reports*, 94(6):535-543, 1979.
- Stoudemire A, Depression in the Medically Ill, in Cavenar, J.O. (ed), *Psychiatry*, Vol. 2, New York:Lippincott, 1985, 1-8.
- Stoudemire A, Frank R, Kamlet M, Hedemark N, Depression, in Robert W. Amler and H. Bruce Dull (eds.) *Closing the Gap: The Burden of Unnecessary Illness*, New York:Oxford University Press, pp. 65-71, 1987.
- Stoudemire, A, Frank R, Kamlet M, Hedemark N, Blazer D, The Economic Burden of Depression, *General Hospital Psychiatry*, 8:387-94, 1986.
- Torrance GW, Sackett DL, Thomas WH, Utility Maximization Model for Program Evaluation: A Demonstration Application, in Robert L. Berg, ed., *Health Status Indexes*, Chicago:Hospital Research and Educational Trust, 1973.
- Torrance GW, Measurement of Health State Utilities for Economic Appraisal: A Review, *Journal of Health Economics* 5(1):1-30, 1986.
- Torrance GW, Boyle MH, Horwood SP, Application of Multiattribute Utility Theory to Measure Social Preferences for Health States, *Operations Research* 30(6):1043-67, 1982.
- Tversky A, Kahneman D, Judgment Under Uncertainty: Heuristics and Biases, *Science*, 185:1124-31, 1974.
- Tversky A, Kahneman D, The Framing of Decisions and the Psychology of Choice, *Science*, 211:453-8.
- VonNeumann J, Morgenstern O, *Theory of Games and Economic Behavior*, Second ed., Princeton, NJ:Princeton University Press, 1947.
- Warner, KE, Systematic Analysis: Cost-Benefit and Cost-effectiveness Techniques, in B. Bloom and S. Berki, eds., *Cost Benefit, Cost-effectiveness, and Other Decision Making Techniques in Health Care Resource Allocation*, 1983.
- Weinstein M, Fineberg H, *Clinical Decision Analysis*, Philadelphia: W.B. Saunders, 1980.

Weinstein MC, Stason WB, Foundations of Cost-effectiveness Analysis for Health and Medical Practices, *New England Journal of Medicine*, 296(13):716-21, 1977.

Weinstein MC, Economic Assessments of Medical Practices and Technologies, *Medical Decision Making* 1(4):309-30, 1981.

Weismann MM, Myers JK, Rates and Risks of Depressive Symptoms in a United States Urban Community. *Acta Psychiatrica Scandinavica*, 57:219-31, 1978.

Weismann MM, The Psychological Treatment of Depression: Update of Clinical Trials, in Williams, J.B.W., and Spitzer, R.L. (eds.), *Psychotherapy Research: Where Are We and Where Should We Go?* New York:Guilford Press, pp. 89-105, 1984.

Weismann MM, The Psychological Treatment of Depression: Evidence for the Efficacy of Psychotherapy Alone, in Comparison with, and in Combination with Pharmacotherapy. *Archives of General Psychiatry*, 36:1216-69, 1979.

Wells KB, Stewart A, Hayes RD, Bumam WR, Daniels M, Berry S, Greenfield S, Ware J, The Functioning and Well-being of Depressed Patients: Results from the Medical Outcomes Study, *Journal of the American Medical Association*, 262(7):914-19, 1989.











